322 J. Opt. Soc. Am. A/Vol. 2, No. 2/February 1985

Model of human visual-motion sensing

Andrew B. Watson and Albert J. Ahumada, Jr.

Perception and Cognition Group, NASA Ames Research Center, Moffett Field, California 94035

Received August 15, 1984; Accepted October 22, 1984

We propose a model of how humans sense the velocity of moving images. The model exploits constraints provided by human psychophysics, notably that motion-sensing elements appear tuned for two-dimensional spatial frequency, and by the frequency spectrum of a moving image, namely, that its support lies in the plane in which the temporal frequency equals the dot product of the spatial frequency and the image velocity. The first stage of the model is a set of spatial-frequency-tuned, direction-selective linear sensors. The temporal frequency of the response of each sensor is shown to encode the component of the image velocity in the sensor direction. At the second stage, these components are resolved in order to measure the velocity of image motion at each of a number of spatial locations and spatial frequencies. The model has been applied to several illustrative examples, including apparent motion, coherent gratings, and natural image sequences. The model agrees qualitatively with human perception.

INTRODUCTION

Visual motion perception serves many roles, among them the perception of depth, the segregation of objects, the control of some eye movements, and the estimation of the motion of objects in the world. Although it is clearly important, the visual-motion sense does not have a widely agreed on principle of operation. One view is that the brain detects a physical property of the stimulus, which we may call movement, sometimes also called optical flow.[1] When this property is detected, the observer has a sensation of visual motion, which may be attached to some visual object. Occasionally, motion is sensed in stimuli that do not contain motion, in which case we speak of apparent motion. An alternative and we believe correct view is that movement is not a physical property of the stimulus. Like beauty and color, motion is in the eye of the beholder. Images do not move but rather change in many and various ways. Some of these changes give rise to an impression of movement. In short, all movement is apparent.

Both of the preceding points of view agree that motion perception is a process that examines the physical stimulus and controls perceptions, judgments, and actions related to the motion of objects or their projections. We adopt the term *motion sensor* for a mechanism at the front end of this process. It begins with the physical stimulus and ends with some explicit representation of quantities related to motion of either the image or the world.

From the optical-flow point of view, the job of the motion sensor is to measure the motion in the stimulus. A model of the sensor is easily constructed, since it is given directly by the physical definition of the quantity to be detected. But, if motion is not a physical property of the stimulus, what is the function of the sensor? It is to sense physical features of the stimulus that are likely (but not certain) to be informative about the motion of objects in the world. The latter case is much more difficult for vision theory. There are many candidate features, and, to model the sensor, we must select a particular set. There are two ways of making this selection. The first is to appeal to invariants in the imagery projected by objects in motion.[2–4] The second approach is to identify a set of features that make sense of the known properties of human vision.[5] In this paper we make use of both approaches.

The plan of this paper is as follows: Section 1 summarizes some properties of human motion perception. In Section 2, we examine the frequency representation of a moving image and propose a set of features to be used by the motion sensor. In Sections 3 and 4, we develop our model of the human motion sensor. An implementation of the model is described in Section 5, and in Section 6 we show some preliminary simulations. In Section 7, we make some concluding comments and suggest how the model may be used to analyze and direct psychophysical, physiological, and computational research.

1. PROPERTIES OF HUMAN MOTION PERCEPTION

In this section, we note some of the important properties of human motion perception. These will guide and motivate the construction of our motion sensor.

A. Stimulus for Motion

The stimulus for visual motion is primarily the distribution of light over time and over two spatial dimensions in a plane before the eyes. Beyond this point, what happens to the signal is sufficiently obscure that a model must be explicit. It is unreasonable for a model to begin with quantities such as edges or optical flow without being explicit about how these quantities are to be computed, or demonstrating that they could be computed, or at least acknowledging that this part of the problem is being sidestepped. In the model developed here, we begin with a distribution of contrast over dimensions of x, y, and t. A digital version of this distribution can be obtained by sampling the output of a video camera.

B. Things Appear to Move with Direction and Speed

Everyday visual experience shows that humans see things move and can with some accuracy judge how fast and in what direction they move. Thus the representation of the image produced by the model should include an explicit assignment of velocity. Whatever other details of the visual process we

0740-3232/85/020322-21$02.00

may try to simulate, our model will have little value if it fails to meet this first test.

C. Perceived Motion Is Local

We are capable of seeing different motions in different parts of an image, such as when two objects move relative to each other. The motion sensor must therefore assign different velocities to different local regions of the visual field. The local assignment of velocity implies that each sensor must respond selectively to a local region of the visual field. There are few data about how local these operations must be, and this is a promising area for further research.

D. Perceived Motion Is Specific to Spatial Frequency

There are several pieces of evidence that show that individual motion sensors respond selectively to different bands of spatial frequency. Perhaps the most compelling is an experiment by Adelson and Movshon[6] in which they superimposed moving sinusoidal gratings of different orientations. When the two gratings were similar in spatial frequency, the percept was of a single coherent plaid pattern with an apparent velocity equal to that of the nodes of the pattern. When the spatial frequencies differed by an octave or so, the two gratings appeared to move independently past each other in their own separate directions.

Another result in this vein is the observation that apparent motion is spatial-frequency specific.[7] Two small patches of spatial grating were presented briefly at nearby positions and again briefly with their positions exchanged. This crossed phi sets up a competition between the perception of patches that remain in static positions and the perception of patches that move and exchange positions. When the spatial frequencies of the two patches are the same, there is no apparent movement. This is not surprising, since nothing has changed from the first presentation to the second. However, when the two spatial frequencies differ by an octave or so, the patches appear to move across each other and exchange positions. If apparent motion were not spatial-frequency specific, there would be no reason for the patches to move. It is not the case that any difference between the two patches suffices to induce apparent motion. For example, if the patches are of the same spatial frequency but different orientation, there is no apparent motion between the patches.

Numerous experiments have shown that the mechanisms that detect stationary contrast are selective for spatial frequency.[8–10] This also appears to be so for moving-contrast patterns.[11] It has also been shown that the mechanisms that detect moving contrast are direction selective (see Subsection 1.G) and signal the direction of motion (at least to within 90 deg) (see Subsection 1.H), properties that are symptomatic of a motion sensor. Taken together, these results suggest that the detectors of moving contrast are spatial-frequency specific.

To summarize, the motion sensor is selective for spatial frequency and may assign different velocities to different spatial-frequency bands. The computation of velocity appears to be carried out independently at a number of different spatial scales.

E. Brief Exposures

Brief exposures to moving stimuli are sufficient to produce vivid motion sensations and accurate motion judgments. For example, at threshold contrast, observers correctly judge the direction (left or right) of a moving grating when the exposure duration is only 410 msec.[12] Increasing the duration by a factor of 5 has little effect.[13] Above threshold, a 5% difference in the speed of a moving line can be discriminated for exposures as short as 200 msec.[14] Evidently, the spatiotemporal feature extracted by the human motion sensor must be brief. Put another way, the sensor must integrate over only a brief interval of time.

F. Adaptation to Motion

Prolonged viewing of a pattern that moves in one direction can produce large alterations in the visibility of subsequently viewed stimuli. These effects are of three general kinds, depending on the type of test stimulus viewed after adaptation. If the test stimulus is in fact stationary, it may appear to move in a direction opposite to the adapting pattern.[15] This is the well-known motion aftereffect. If the test stimulus moves in a direction similar to that of the adapting pattern, its apparent direction may be repulsed away from the adapting direction by as much as 10 deg.[16] And, finally, contrast thresholds are raised much more for test stimuli that move in the adapting direction than for those that move in the opposite direction.[17]

All these effects suggest the existence of separate, adaptable, direction-selective sensors. They further suggest that the final estimate of velocity is the result of computations across a population of sensors selective for different directions.

G. Subthreshold Summation

The threshold contrast for a sinusoidal grating moving to the right is almost unaffected by a grating of equal contrast moving to the left.[12,18,19] This is so in spite of the fact that the sum of two gratings that move in opposite directions has twice the peak contrast of either grating alone. The result indicates that the mechanisms that detect moving patterns are direction selective, in the sense that they are sensitive to less than the full 360-deg range of movement directions. It also indicates that, at least near threshold, the detectors of opposite directions are independent of one another.

This result is observed only when the image velocity is greater than about 1 deg/sec.[12,19] Below this point, considerable summation is observed between gratings that move in opposite directions, consistent with detection by nondirection-selective mechanisms. This agrees with other evidence that below this point a separate, nondirection-selective, *static* system is more sensitive (see Subsection 1.K).

When two gratings of the same spatial frequency u moving in opposite directions at speed r are added together, the result is a stationary grating whose contrast varies sinusoidally in time (flickers) with a temporal frequency of $w = ru$ and whose contrast is twice that of either moving component. The data cited indicated that at low velocities the thresholds for moving and flickering gratings of the same spatial and temporal frequency are about equal, while at higher velocities sensitivity is almost twice as great for the moving grating. Over the full range of velocities, therefore, sensitivities to these two patterns are equal to within a factor of 2.

H. Direction Discrimination at Threshold

Observers are able to tell which way things are going at the threshold of vision. More precisely, the contrast threshold

for discriminating whether a grating moves to the right or to the left is the same as the contrast threshold for detecting the stimulus.[12] This is consistent with the hypothesis that moving stimuli are detected by sensors that are direction selective and that explicitly assign a direction to the internal representation of the stimulus. In other words, direction is part of the label of the sensor.

This result is found only for images moving at moderate to high velocities (above about 1 deg/sec). At lower velocities, the threshold for judging direction is much higher than that for simple detection. This is not due to the confounding effects of eye movements on slowly moving patterns, as it has been demonstrated under stabilized conditions.[20] This discrepancy between results at low and high velocities may reflect the operation of separate motion and static systems (see Subsection 1.K).

I. Contrast Sensitivity to Moving Patterns

Robson[21] and others[22,23] have measured human sensitivity to flickering sinusoidal gratings with specific spatial and temporal frequencies. The data form a surface as a function of spatial and temporal frequency whose high-frequency boundary is the product of separable spatial and temporal contrast-sensitivity functions.[21] In other words, the upper bounds of sensitivity appear to be set by separate spatial and temporal filtering processes. As noted in Subsection 1.K, contrast thresholds for moving and flickering gratings are always equal to within a factor of 2. Apart from this small factor, then, this spatiotemporal contrast-sensitivity surface determines sensitivity limits for moving gratings. Direct measurements of sensitivity to moving gratings indicate that this is so.[24,25]

One consequence of the shape of the spatiotemporal contrast-sensitivity surface is that, as the image moves faster, the highest spatial frequency that can be seen declines. As speed increases, the temporal frequency (the product of speed and spatial frequency) increases most rapidly for the highest spatial frequencies. We have already observed that motion sensors must be selective for different bands of spatial frequency. The above result implies that low-spatial-frequency motion sensors will respond to higher velocities than will high-spatial-frequency motion sensors.

J. Apparent Motion

If a spatial target is presented at one location, then extinguished, and a brief time later presented again at a nearby location, it may appear to move between the two points. A summary of the extensive classical literature on this subject is provided by Kolers.[26] The effect is more pronounced when the sequence contains not just two but many presentations along a path.[27] When the time and space intervals between presentations are brief enough (as in film and video), the sequence of static presentations may be indistinguishable from continuous motion. This limiting result is probably a consequence of the known spatial and temporal filtering actions of the visual system and does not necessarily reveal anything about the motion-sensing system *per se*.[29,30]

Apparent motion is important for three reasons. First, it is a good example of a stimulus without a well-defined optical flow, which may yet give rise to unambiguous perceived motion. Second, it imposes another constraint on our model. It too must respond to apparent-motion stimuli and should indicate the same speed and direction as perceived by the observer. Third, it suggests a filter structure for our sensor. The stimulus for apparent motion may be regarded as a sampled version of a corresponding continuous motion. The sampled stimulus differs from the continuous stimulus by the addition of certain spatial and temporal frequencies. The apparent similarity or identity of the two stimuli suggests that the additional frequencies have been filtered out. This leads us to seek a filter structure for our motion sensor and to investigate what form it should have in the space–time frequency domain.[5,7,30] This investigation is carried out in Section 2.

K. Motion and Static Systems

There is some evidence that the visual system contains two separate systems to process and represent moving and static imagery, respectively.[17] The systems are thought to be distinct in their spatiotemporal sensitivity, the motion system being more responsive to rapidly moving patterns (low spatial and high temporal frequencies) and the static system more responsive to slowly moving patterns (high spatial and low temporal frequencies). In addition, the motion system is thought to be direction selective, in the sense that image components moving in opposite directions are sensed by separate mechanisms. The systems are also thought to differ in the image representations that they provide to later stages of visual processing: The motion system assigns velocity (or at least direction) values to components in an image; the static system does not.

The evidence on this point is modest but favorable. As noted above (see Subsection 1.G), contrast summation is direction selective only above about 1 deg/sec.[12] Furthermore, direction is judged correctly at threshold only above about 1 deg/sec (see Subsection 1.H).[12,13,20] Discrimination of spatial frequency at detection threshold appears to be much worse when the temporal frequency is suited to the motion system than when it is suited to the static system.[31]

L. Inhibition between Directions

There is some evidence for inhibitory interactions between the sensors for opposite directions.[32,33] Levinson and Sekuler[32] showed that adaptation to the sum of leftward- and rightward-moving gratings produced less threshold elevation for a rightward-moving grating than did adaptation to a rightward-moving grating alone. In addition, it is difficult to construct explanations of the motion aftereffect that do not involve inhibition between the sensors for opposite directions.

M. Speed Discrimination

Some of the evidence cited indicates that motion sensors are selective for spatial frequency and for direction. Another dimension in which they might be selective is speed. However, observers are quite poor at discriminating the speed of moving (or flickering) gratings at detection threshold.[31,34] Quite low speeds can be discriminated from quite high speeds, but no finer performance is possible. This suggests that the sensors are not particularly selective for speed or that the sensors that detect moving images do not explicitly assign a speed value. Above threshold, however, the system is exquisitely sensitive to variations in speed, reliably detecting variations on the order of 5%.[14] This discrepancy between

threshold and suprathreshold discrimination is a common feature of human perception (see, for example, the comparable case in discrimination of spatial frequency[31,35]) and may be explained by a two-stage model. The second stage, which becomes effective only above threshold, combines the responses of first-stage mechanisms and provides finer performance. Our motion sensor will take this two-stage form.

2. FREQUENCY COMPOSITION OF MOVING IMAGES

Section 1 described some properties of human motion perception that will guide the construction of our motion sensor. One of those properties was that motion analysis appears to be done in parallel within a number of separate bands of spatial frequency and that the elementary features into which images are decomposed are patches of oriented sinusoid. This leads us to consider how the frequencies that make up an image behave as the image is moved.[5,28–30,36] We restrict our attention to one particular case: that of an otherwise-unchanging image undergoing translation at a constant velocity. This is one of the rare cases in which an unambiguous image velocity exists. It should be noted that other, more-complex motions can be constructed by piecing together spatial and temporal segments of this simple kind.

A. Contrast Distribution of a Moving Image

An arbitrary monochromatic space–time image can be represented by a function $c(x, y, t)$ defined over some interval, which specifies the contrast at each point x, y and time t. Contrast is defined as the luminance divided by the average luminance over the interval. Although not strictly necessary, it is convenient to begin this discussion with an image that is static, that is,

$$c_0(x, y, t) = c_0(x, y, 0) \qquad \text{for all } t. \tag{1}$$

Define the image velocity as the vector $\mathbf{r}$ with horizontal and vertical speed components r_x and r_y, or, in polar terms, a speed r and direction θ, where $r_x = r\cos\theta$ and $r_y = r\sin\theta$. If the previously static image translates at constant velocity $\mathbf{r}$, the distribution becomes

$$c_{\mathbf{r}}(x, y, t) = c(x - r_x t, y - r_y t, t). \tag{2}$$

B. Fourier Transform of a Moving Image

Given an arbitrary space–time image $c(x, y, t)$, and its Fourier transform $\tilde{c}(u, v, w)$, we seek a general expression for the transform of this image as it undergoes translation at a constant velocity $\mathbf{r}$.

1. *Two-Dimensional Case*

For simplicity, we develop the two-dimensional (2D) case of $c(x, t)$. This would be adequate to describe images without variation in the vertical dimension, for example, a vertical line or grating. We adopt the following vector notation:

$$\mathbf{a} = \begin{pmatrix} x \\ t \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} u \\ w \end{pmatrix}, \tag{3}$$

where u and w are the spatial- and temporal-frequency variables corresponding to x and t, respectively. Then a function and its 2D Fourier transform can be written as

$$c(\mathbf{a}) \rightarrow_2 \tilde{c}(\mathbf{b}), \tag{4}$$

where $\rightarrow_2$ indicates the 2D Fourier transform. Translation of the image can be represented as a coordinate transformation in which the x coordinate increases linearly with time. Let r be the speed of horizontal motion. If we write $\mathbf{a}' = \begin{pmatrix} x' \\ t' \end{pmatrix}$ for the transformed coordinates, then

$$\mathbf{a}' = \begin{pmatrix} x - rt \\ t \end{pmatrix} = \mathbf{A}\,\mathbf{a}, \qquad \mathbf{A} = \begin{bmatrix} 1 & -r \\ 0 & 1 \end{bmatrix}. \tag{5}$$

From the general expression for the transform following an affine coordinate transformation (any combination of scaling, rotation, and translation),[37] we find that

$$c(\mathbf{a}') \rightarrow_2 \tilde{c}[(\mathbf{A}^{-1})^T\mathbf{b}], \tag{6}$$

where the superscript $[\ ]^T$ indicates the matrix transpose operation. Since

$$(\mathbf{A}^{-1})^T = \begin{bmatrix} 1 & 0 \\ r & 1 \end{bmatrix}, \tag{7}$$

we end up with

$$c(x - rt, t) \rightarrow_2 \tilde{c}(u, w + ru). \tag{8}$$

Graphically this means that moving the image shears its spectrum in the w dimension. Spatial frequencies are not changed, but all temporal frequencies are shifted by minus the product of the speed and the spatial frequency, $-ru$. This result is easiest to picture for an image that was static in time before the introduction of motion, as in Fig. 1. The spectrum of the static image lies entirely along the u axis (all signal energy is at 0 Hz). When the image moves, the spectrum is sheared as described above, so that the result lies along a line through the origin with a slope of $-r$.

2. *Three-Dimensional Case*

The preceding can be generalized to three dimensions. If the velocity is $\mathbf{r} = (r_x, r_y)$, the moving image and its transform are then

$$c(x - r_x t, y - r_y t, t) \rightarrow_3 \tilde{c}(u, v, w + r_x u + r_y v). \tag{9}$$

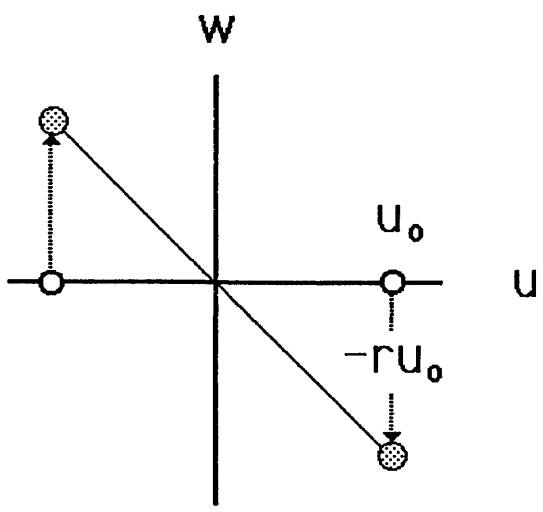


Fig. 1. The effect of motion on the Fourier transform of a 2D space–time image. The effect is shown for a single representative component of spatial frequency u_0. The open circles show the location of the components of the static image. The dotted circles show the locations when the image moves with speed r. Each transform point is sheared in the w dimension by an amount $-ru$, as indicated by the arrows.

Each temporal frequency is shifted by an amount $-(r_x u + r_y v)$. If the original image was static ($w = 0$), then the new temporal frequency will be equal to this quantity. Since it will play an important role in subsequent developments, we note that this quantity is the dot product of the two-dimensional spatial frequency vector $\mathbf{f} = (u, v)$ and the image-velocity vector $\mathbf{r} = (r_x, r_y)$. Three possible expressions for the temporal frequency of each component of the spectrum of an image in motion are then

$$w = -\mathbf{r} \cdot \mathbf{f} = -(r_x u + r_y v) = -rf \cos(\theta - \alpha), \quad (10)$$

where θ is the direction of image motion and α is the orientation of the spatial-frequency component. This quantity can be seen to be the product of the spatial frequency f and *the component of the velocity in the direction equal to the orientation of the grating.*[38] Note that the two-dimensional result described earlier is a degenerate case of this general outcome.

Geometrically, image motion changes the static-image transform, which lies in the u, v plane, into a spectrum that lies in an oblique plane whose intersection with the u axis has a slope of $-r_y$ and whose intersection with the v axis has a slope of $-r_x$ (the dihedral angle of the plane will be $\tan^{-1} r$). To illustrate, consider a stationary sinusoidal grating of frequency f and orientation α. As illustrated in Fig. 2, its transform is a pair of three-dimensional (3D) impulses at u_0, v_0, 0 and $-u_0$, $-v_0$, 0, where $u_0 = f \cos \alpha$ and $v_0 = f \sin \alpha$. If the grating moves in direction θ at speed r, applying Eq. (10) shows that the spatial frequencies will not change, but the temporal frequencies will shift to minus and plus $rf \cos(\theta - \alpha)$, respectively. These two points lie at opposite ends of a line through the origin.

In summary, the spectrum of a stationary image lies in the u, v plane. When the image moves, the transform shears into an oblique plane through the origin. The orientation of this plane indicates the speed and direction of motion.

The preceding discussion shows that it is possible to associate energy in particular regions in spatiotemporal frequency space (for example, the one pictured by the dotted circles in Fig. 2), with particular image-velocity components. By filtering specific regions, it is therefore possible to detect image-velocity components of particular values. This observation forms the basis for the development of our scalar motion sensor. However, as noted below, a single image-velocity component is ambiguous. The ambiguity must be resolved by combining several measurements of different components of the image velocity.

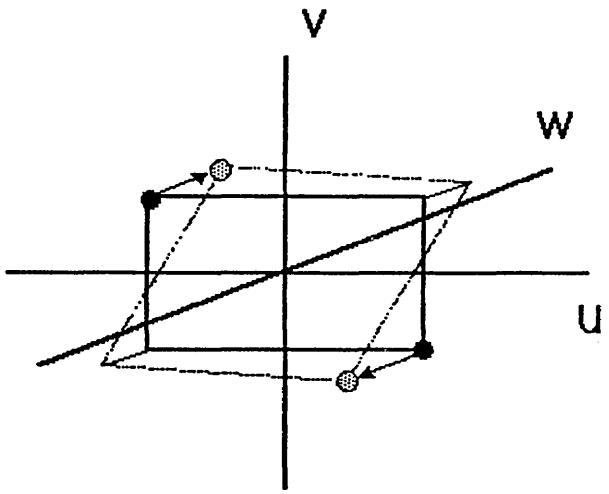


Fig. 2. The effect of motion on the transform of a 3D space–time image. The effect is shown both for a plane representing the full spectrum and for a single representative component of spatial frequency f and orientation α. The solid plane and filled circles show the location of the spectrum of the static image defined by $w = 0$. Motion at velocity $\mathbf{r}$ shears the spectrum into the plane $w = \mathbf{r} \cdot \mathbf{b}$, as shown by the dashed–dotted plane and dotted circles. The arrows indicate the displacement of a single spatial-frequency component.

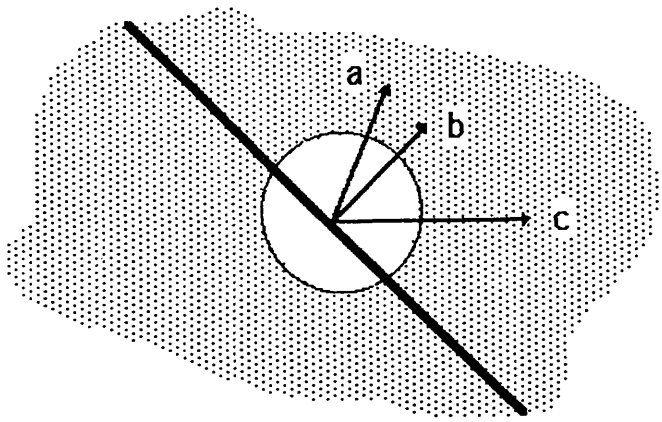


Fig. 3. An example of direction ambiguity. The motion of the contour seen through the aperture is consistent with any of the velocities a or b or c (after Marr and Ullman). Note that all possible velocities have the same velocity component orthogonal to the contour.[39]

C. Direction Ambiguity

When a straight contour moves behind an aperture (Fig. 3), its direction is ambiguous. The stimulus is consistent with a range of velocities whose directions cover 180 deg. This has been called the *aperture problem.*[39] In fact, the aperture is not critical, for the same ambiguity will result from a pattern of infinite extent, provided that all its frequency components have the same orientation. The aperture plays its role by limiting the visible portion of a pattern to a region within which only a single orientation is present.

This ambiguity is easily seen in the frequency-space diagram of Fig. 2. The plane, which contains spatial-frequency components at all orientations, uniquely defines an image velocity of r, θ. However, the solid circles, which represent a single component with just one orientation, lie within many possible planes and are thus consistent with many possible velocities. Similarly, a moving image whose spatial frequencies all have the same orientation has a spectrum that lies along a straight line through the origin. This line is contained within an infinite number of planes and is thus consistent with an infinite number of velocities. Specifically, if the orientation of the image components is α, and the speed in direction α is r_α, then the transform (and image) corresponds to any velocity r_θ, θ such that

$$r_\theta = r_\alpha \cos(\theta - \alpha). \quad (11)$$

This ambiguity has traditionally been a problem for motion sensors that operate on individual oriented components in the image. The problem of how these ambiguous oriented components are resolved into a single unambiguous velocity estimate is dealt with in Section 4.

D. Summary

In Section 1, we observed that moving images are sensed in terms of their oriented 2D spatial-frequency components. Here, we have shown that the spectrum of a moving 2D spatial-frequency component occupies two regions at opposite ends of a line through the origin. A sensor designed to filter such antisymmetric regions in frequency space should therefore sense these moving components. This sensor is constructed in Section 3. However, as noted in Subsection 2.C, individual oriented components are ambiguous as to direction, in that each indicates only one scalar component of the image velocity. Therefore additional steps must be taken to remove this ambiguity. We take up this point in Section 4.

3. THE SCALAR MOTION SENSOR

We construct our model in two stages. The first stage is an array of linear sensors that are selective for 2D location, 2D spatial frequency, and direction. Because, at any moment in time, each produces only a single number rather than the vector required to specify velocity, we call it a scalar motion sensor. The second stage, discussed in Section 4, is called the vector motion sensor.

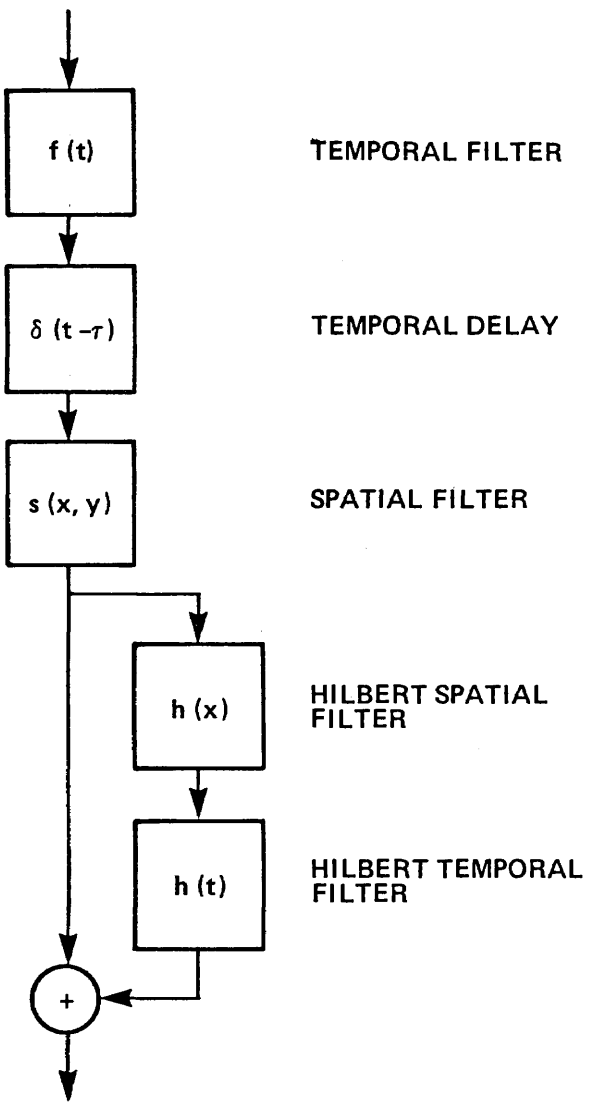


Fig. 4. Mathematical structure of the scalar motion sensor.

The scalar sensor can be described in two ways. One is a mathematical derivation; the other is a physical or physiological description, without reference to mathematical detail. We begin with the mathematical treatment.

The overall mathematical structure of the scalar sensor is illustrated in Fig. 4. It is an assembly of spatial and temporal filters each of which has simple properties. In assembling these components, two points should be remembered:

(1) When several filters are in series, their order is irrelevant. This is formally expressed in the commutative property of convolution, $f * g = g * f$. Consequently, the order in which we introduce components is not necessarily the order in which their physical analogs are connected.

(2) The filter that we are constructing operates on the three dimensions x, y, t or in frequency space u, v, w. A separable filter is one in which the impulse response is the product of one-dimensional (1D) components. Separability permits certain simplifications; for example, the 2D or 3D transform of a separable function is the product of the separate 1D transforms, and the 2D or 3D convolution of separable functions is the product of their separate 1D convolutions. For these reasons, we will deal separately with temporal, horizontal, and vertical filters, which should be understood as the 1D components of 3D filters. For example, $f(t)$ is the time component of a 3D impulse response $f(t)\delta(x)\delta(y)$.

A. Basic Temporal Filter

We introduce a temporal filter with impulse response $f(t)$ and transfer function $\tilde{f}(w)$. We have tentatively identified this function with the temporal sensitivity of the human observer to relatively low spatial frequencies.[21–23,40,41] A useful analytic approximation to these data is provided by the impulse response

$$f(t) = \xi[f_1(t) - \zeta f_2(t)], \tag{12}$$

where

$$f_i(t) = \frac{u(t)}{\tau_i(n_i - 1)!}(t/\tau_i)^{n_i-1}e^{-t/\tau_i} \tag{13}$$

and $u(t)$ is the unit step function. The function f_i is an n-stage low-pass filter, essentially as used by Fourtes and Hodgkin.[42] This analytic form of $f(t)$ has been used elsewhere to model temporal contrast-sensitivity data.[41,43] The filter transfer function is given by the Fourier transform of the impulse response,

$$\tilde{f}(w) = \xi[f_1(w) - \zeta\tilde{f}_2(w)], \tag{14}$$

where

$$\tilde{f}_i(w) = (i2\pi w\tau_i + 1)^{-n_i}. \tag{15}$$

We have chosen parameters ($\zeta = 0.9$, $\tau_1 = 0.004$, $\tau_2 = 0.0053$, $n_1 = 9$, $n_2 = 10$) for this filter that fit the data of Robson.[21] Impulse response, amplitude response, and phase response of the filter are shown in Fig. 5. The operation of the sensor is not critically dependent on the particular version of $f(t)$ selected. An alternative choice might be the temporal impulse responses of individual visual cells, but these appear to be quite varied.

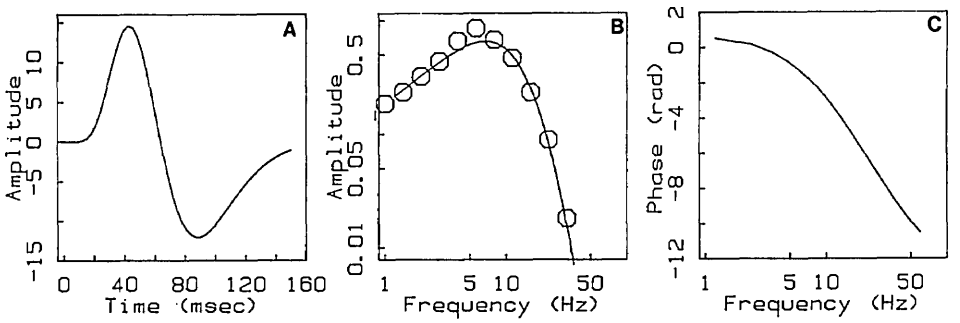


Fig. 5. A, Impulse response; B, amplitude response; C, phase response of the basic temporal filter. The symbols in the center panel are contrast-sensitivity measurements made by Robson[21] for a 0.5-cycle/deg grating.

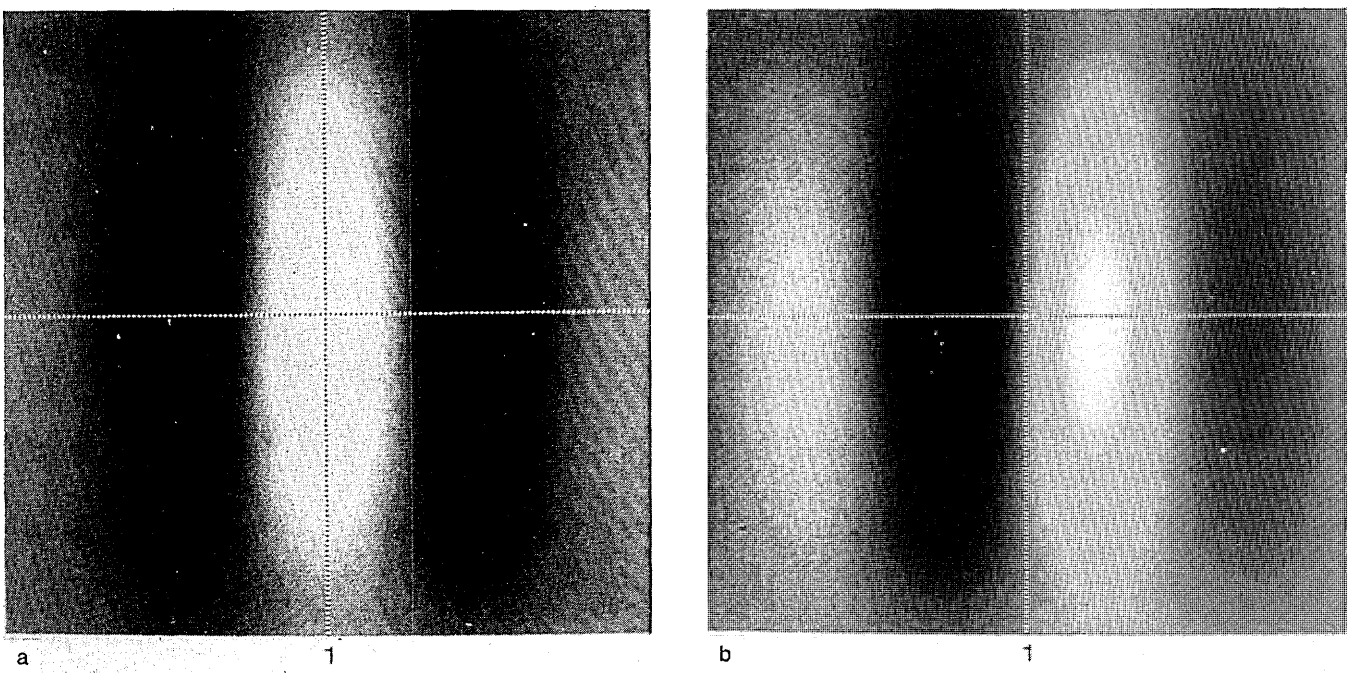


Fig. 6. Spatial impulse responses of a, main and b, quadrature paths. The spatial impulse response of the basic spatial filter is equivalent to that of the main path, a. The gray background indicates a value of 0.

B. Basic Spatial Filter

We next introduce a spatial filter with impulse response $g(x, y)$ that is separable into horizontal and vertical components

$$g(x, y) = a(x)b(y), \tag{16}$$

$$a(x) = \exp[-(x/\lambda)^2]\cos(2\pi u_s x), \tag{17}$$

$$b(y) = \exp[-(y/\lambda)^2]. \tag{18}$$

This impulse response is pictured in Fig. 6. It is the product of a cosine and a Gaussian in the horizontal dimension and is a Gaussian in the vertical dimension. The frequency of the cosine is u_s, and the Gaussians both have spreads of λ. This function, sometimes called a 2D Gabor function, is a good approximation to the receptive-field weighting function of so-called simple visual cortical cells in cat and monkey[44] and presumably in the human. It has also been used as a model of the basic spatial operator in theories of human spatial vision.[45,46]

The transform is

$$\tilde{g}(u, v) = \tilde{a}(u)\tilde{b}(v), \tag{19}$$

$$\tilde{a}(u) = \sqrt{\pi}\lambda/2\,(\exp\{-[\pi\lambda(u - u_s)]^2\} + \exp\{-[\pi\lambda(u + u_s)]^2\}), \tag{20}$$

$$\tilde{b}(v) = \sqrt{\pi}\lambda \exp[-(\pi\lambda v)^2]. \tag{21}$$

C. Size of the Spatial Sensor

The parameter λ governs the size of the Gaussian component of the spatial impulse response. Here, we set it equal to a constant ρ times the period of the sinusoid,

$$\lambda = \rho/f. \tag{22}$$

Thus the size of the sensor scales inversely with its spatial frequency. This has the consequence of fixing the log-frequency bandwidth of the sensor. We assume that $\rho = 3\sqrt{\ln 2}/\pi = 0.795$, in which case this fixed log bandwidth is one octave. This particular bandwidth is chosen to be consistent with available psychophysical and electrophysiological evidence.

With the introduction of the basic spatial and temporal filters we have traversed the first two boxes in Fig. 4. The resulting transfer function will be the product of the temporal and spatial transfer functions. Since each of these has two

lobes, one on either side of the frequency origin, their product will have four lobes rather than the two antisymmetric lobes required by a direction-selective sensor (see Fig. 1). Our next step then is to null two of the four lobes of the filter transfer function and thus make the sensor direction selective. We do this by means of a Hilbert filter.

D. Hilbert Transform

We digress to introduce the Hilbert transform, which is dealt with in most standard treatments of linear systems theory but is rarely seen in the vision literature. The Hilbert transform of a function $g(x)$ is given by $h(x) * g(x)$, where

$$h(x) = -1/\pi x. \tag{23}$$

The Hilbert transform is a linear filter with an impulse response that is an inverted hyperbola. The transfer function of the filter is the Fourier transform of $h(t)$:

$$\tilde{h}(u) = i\,\mathrm{sgn}(u). \tag{24}$$

This filter, pictured in Fig. 7, has the remarkable property of unit gain at all frequencies combined with a constant phase lag of $\pi/2$ for positive frequencies and $-\pi/2$ for negative frequencies. It converts odd functions into evens, and evens into odds. Two functions that are Hilbert transforms of each other are said to form a quadrature pair.

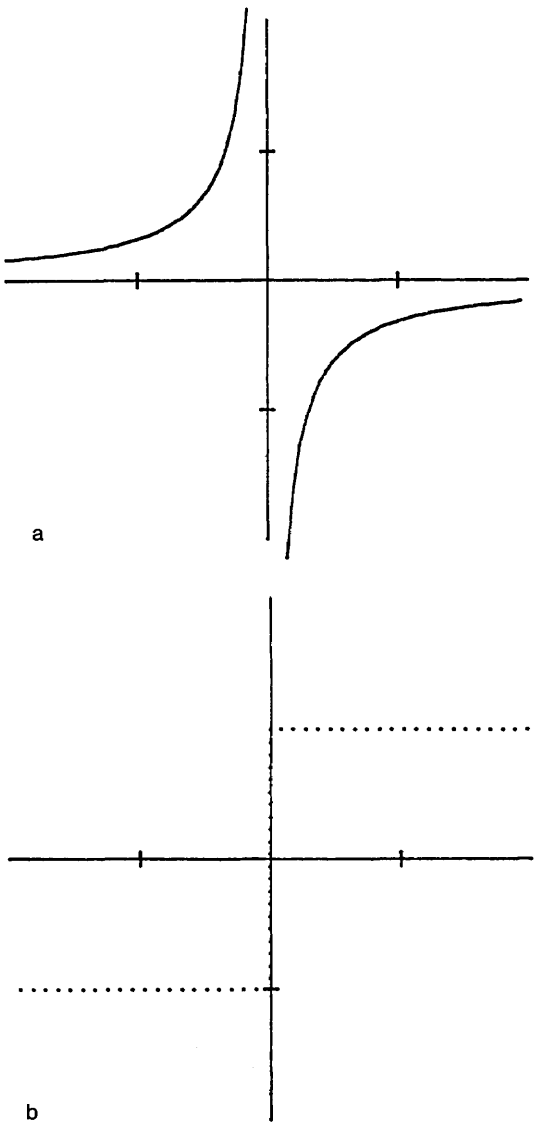


Fig. 7. The Hilbert filter. a, Impulse response. b, Transfer function. The dotted line indicates an imaginary value.

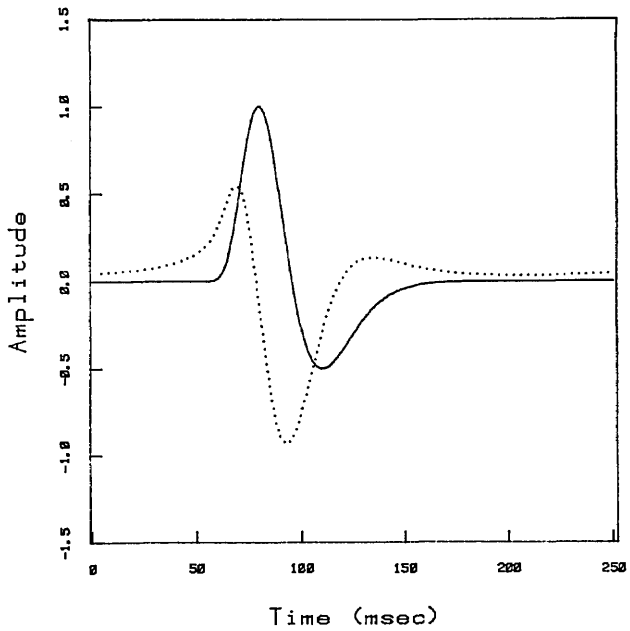


Fig. 8. Temporal impulse responses of main (solid line) and quadrature (dotted line) paths of the scalar motion sensor.

E. Time Delay

The next element in the sensor is a time delay of τ. This can be represented by convolution by a delayed impulse $\delta(t - \tau)$. The delay has a transfer function of $\exp(-i2\pi w\tau)$. This delay is introduced to ensure that the introduction of the Hilbert filter, described below, does not result in a noncausal temporal impulse response.

F. Main and Quadrature Paths

At this point, the signal path branches into what we will call main and quadrature paths. The quadrature path is subject to a Hilbert transform in each of the dimensions of t and x.

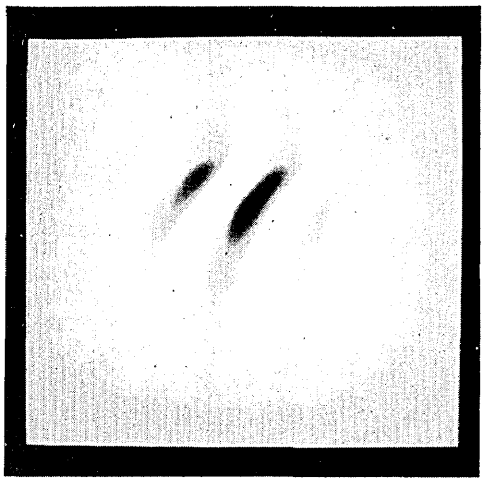
G. Hilbert Temporal Filter

In the quadrature path, the signal passes through a temporal Hilbert filter, so that the temporal impulse responses of main and quadrature paths are now $f(t - \tau)$ and $f(t - \tau) * h(t)$, respectively. These are illustrated in Fig. 8. The Hilbert impulse response, when viewed as a function of time, is neither causal nor physically realizable. However, we need only ensure that the result of the transform (the dashed curve in Fig. 8) is causal and realizable, which is easily done through suitable choice of the delay τ. The Hilbert impulse response contains infinite values, but we make use of it in the frequency domain, where it presents no difficulties of calculation.

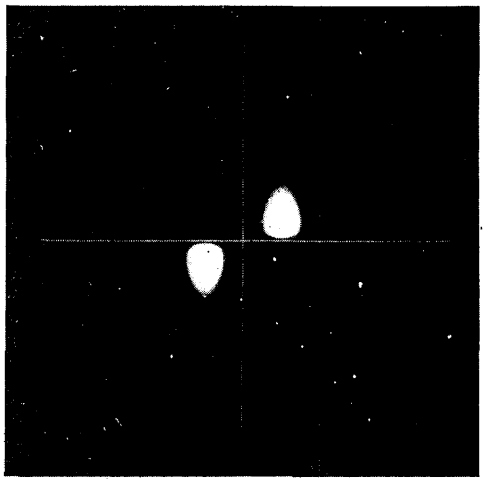
H. Hilbert Spatial Filter

The quadrature path is now subjected to another Hilbert transform, this time in the horizontal spatial domain. The resulting spatial impulse response in the quadrature path is

$$[h(x) * a(x)]b(y) \tag{25}$$



a



b

Fig. 9. a, Impulse response and b, amplitude response of a scalar sensor for leftward motion (direction = 0). In a, the axes are x (horizontal) and t (vertical). In b, they are u and w.

with a Fourier transform

$$i\,\mathrm{sign}(u)\tilde{a}(u)\tilde{b}(v) \tag{26}$$

Equation (20) shows that $\tilde{a}(u)$ is two Gaussians at u_s and $-u_s$ with spreads of $1/(\pi\lambda)$. If, as in the present case, $\lambda > 2/(\pi u_s)$, then each Gaussian will be almost entirely on one side of the origin, in which case the transform will be the sum of two complex Gaussians, one positive at u_s, the other negative at $-u_s$. Taking the inverse transform shows that the resulting impulse response in the quadrature path is simply (approximately) equal to minus a sinusoid multiplied by a Gaussian. Thus the spatial impulse responses of the main and quadrature paths are even and odd (cosine and sine phase) Gabor functions. They are shown in Fig. 6.

I. Sensor Transfer Function

We have applied Hilbert filters of t and x to the quadrature path; hence its transfer function $\tilde{m}_q$ will differ from that of the main path $\tilde{m}_m$ by multiplication by the transfer functions of these two Hilbert filters,

$$\tilde{m}_m(u, v, w) = \tilde{a}(u)\tilde{b}(v)\tilde{f}(w)\exp(-i2\pi w\tau), \tag{27}$$

$$\tilde{m}_q(u, v, w) = -\tilde{m}_m(u, v, w)\mathrm{sign}(u)\mathrm{sign}(w). \tag{28}$$

The final step is to combine the two paths. For a sensor of rightward motion, we add the two paths,

$$\tilde{m}_r(u, v, w) = \tilde{m}_m + \tilde{m}_q = \tilde{a}(u)\tilde{b}(v)\tilde{f}(w) \times \exp(-i2\pi w\tau)[1 - \mathrm{sign}(u)\mathrm{sign}(w)]. \tag{29}$$

The modulus of this transfer function (the amplitude response), and the corresponding impulse response, are shown in Fig. 9. Note that the amplitude response occupies only the second and fourth quadrants of the u, w space. Thus the effect of adding the quadrature path is to null two of the four lobes of the transfer function, so that it occupies only two diagonally opposite quadrants in frequency space. Recall that these are the quadrants occupied by an image moving to the right (Fig. 1). An image moving to the left will occupy the first and third quadrants and will therefore produce no response. The sensor is therefore selective for motion to the right.

J. Generalization to Other Directions

Each scalar sensor has a particular spatial frequency f_s and direction θ_s. These numbers can be compactly represented by a vector $\mathbf{s} = (u_s, v_s)$, where $u_s = f_s \cos\theta_s$ and $v_s = f_s \sin\theta_s$. We will call this the *directed frequency* of the sensor. So far we have developed the case of a sensor for horizontal motion to the right, that is, $\theta_s = 0$. The sensor for an arbitrary direction θ_s is obtained by rotating the x, y space by $-\theta_s$. This means that we are using a new spatial weighting function with orientation θ_s. In the frequency domain, this has the effect of rotating the transform by θ_s, which we accomplish by the coordinate transformation

$$u' = u\cos\theta_s + v\sin\theta_s, \qquad v' = -u\sin\theta_s + v\cos\theta_s. \tag{30}$$

Let $G = \pi\lambda^2/2$. This is the overall gain of the motion sensor. Then the transfer function of a sensor of directed frequency is

$$\tilde{m}_{\mathbf{s}}(\mathbf{f}, w) = G\{\exp[-(\pi\lambda|\mathbf{s} - \mathbf{f}|)^2] + \exp[-(\pi\lambda|\mathbf{s} + \mathbf{f}|)^2]\} \times \tilde{f}(w)\exp(-i2\pi\tau w)[1 - \mathrm{sgn}(\mathbf{s}\cdot\mathbf{f})\mathrm{sgn}(w)]. \tag{31}$$

Note that the first term is a pair of Gaussians at $\mathbf{s}$ and $-\mathbf{s}$. The impulse response and the amplitude response of the scalar sensor for one direction are shown in Fig. 10.

K. Cardinal Form for a Direction-Selective Sensor

It is evident that the essential feature of our direction-selective sensor is that it responds in just two antisymmetric quadrants of frequency space. In the preceding sections we have developed a particular example of such a sensor, but it is worth considering which of its features are essential to meet the above criterion. A canonical form for our linear direction selective sensor (in the $\mathbf{x}$ direction) is

$$a(x)b(y)c(t) * [\delta(x, y, t) + \delta(y)h(x)h(t)], \tag{32}$$

where a, b, and c are arbitrary separable functions of space and time. In other words, the filter consists of the product of separable functions, plus its own Hilbert transform in x and t. The canonical forms for other directions are obtained through suitable coordinate rotations. Thus it is the structure of form (32) that gives the direction selectivity; the functional forms of a, b, and c have been chosen to fit other aspects of human motion perception.

L. Simple Description of the Scalar Motion Sensor

As promised at the start of Section 3, we can now translate the preceding discussion into a simple physiological description of the sensor. We imagine a pair of visual neurons, with spatial receptive fields as pictured in Fig. 9. The functions in this figure now describe the influence of light upon each cell as a function of the distance of the point of light from the center of the receptive field. Second, we interpret the two curves in Fig. 8 as the temporal impulse responses of the two cells. For example, the solid curve in Fig. 8 is the time response of the even cell (Fig. 9a) to a brief pulse input, whereas the dashed curve is the similar response for the odd cell. We then imagine a third cell that simply adds the responses of the odd and the even cells. This third cell will be direction selective and is a physical embodiment of our scalar motion sensor.

M. Response of the Scalar Motion Sensor

The response of the sensor to an arbitrary input is given directly by convolution

$$r(x, y, t) = c(x, y, t) * m(x, y, t) \tag{33}$$

or by way of the convolution theorem

$$r(x, y, t) \rightarrow_3 \tilde{c}(u, v, w)\tilde{m}(u, v, w). \tag{34}$$

The mathematical development of the sensor as a filter results in a response that is itself a 3D function of space and time. Yet when speaking about a sensor as the theoretical analog of a single visual cell we expect its response to be a function of time only. This apparent contradiction is resolved by viewing $r(x, y, t)$ as specifying the response at time t of the sensor centered at x, y.

What is the form of this temporal response? The sensor removes all spatial frequencies save those in the neighborhood of $\mathbf{s}$. As noted in Subsection 2.B.1, the input at $\mathbf{s}$ has a temporal frequency of $\mathbf{s} \cdot \mathbf{r}$. Since the sensor is linear, the output will have the same frequency as the input. Thus the temporal response of the sensor will approximate a sinusoid with a temporal frequency of

$$w_s = \mathbf{s} \cdot \mathbf{r}. \tag{35}$$

Comparing this with Eq. (10), we note the following useful result: *The temporal frequency of the sensor response approximates the component of the image velocity in the sensor direction, multiplied by the spatial frequency of the sensor.* We can think of the temporal frequency as *coding* one component of the image velocity.

N. Selectivity of the Scalar Motion Sensor

Examination of the transfer function of the motion sensor reveals its selectivity for direction, orientation, spatial frequency, temporal frequency, and the like. A simple way of showing this is by considering the response to sinusoidal gratings of various spatial frequencies, orientations, directions and velocities. Since the system is linear, the temporal response to this stimulus will be sinusoidal with the same temporal frequency as the input. It will have an amplitude that can be read directly from the amplitude response given in Eq. (31). For example, consider a 2D sinusoid of spatial frequency $\mathbf{f}$ and velocity $\mathbf{r}$. From Eq. (31) it is evident that the response will be a sinusoid of amplitude

$$G \exp[-(\pi\lambda|\mathbf{s} - \mathbf{f}|)^2]|\tilde{f}(|\mathbf{s}||\mathbf{f}|)|[1 - \text{sign}(\mathbf{s} \cdot \mathbf{r})]. \tag{36}$$

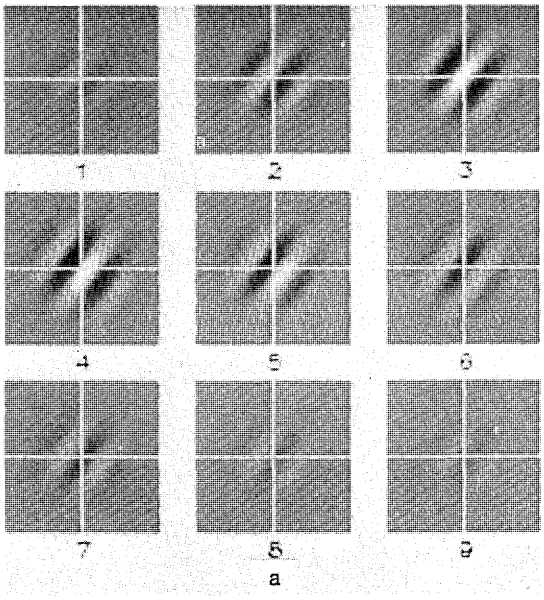
This expression can now be used to illustrate the sensitivity of the sensor to spatial frequency and direction.

1. *Spatial Frequency*

If grating direction matches that of the sensor ($\theta = \theta_s$), then the sensor response amplitude is given by

$$G \exp\{-[\pi\lambda(f - f_s)]^2\}|\tilde{f}(rf)|. \tag{37}$$

When the temporal frequency $w = rf$ is fixed, this amplitude



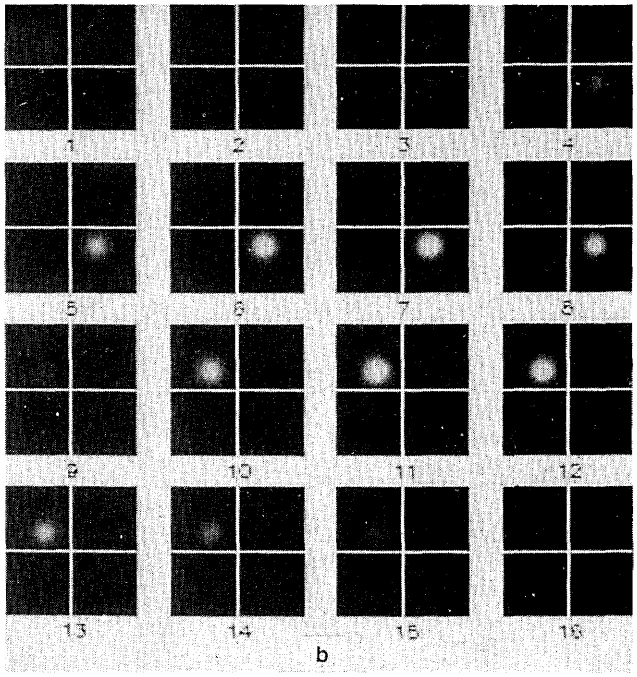


Fig. 10. a, Impulse response and b, amplitude response of a scalar sensor for motion in direction $\theta_s = 324°$. In a, frames show successive time samples of 12.5 msec. In b, frames show successive temporal frequency samples of 5 Hz, with the origin in frame 9.

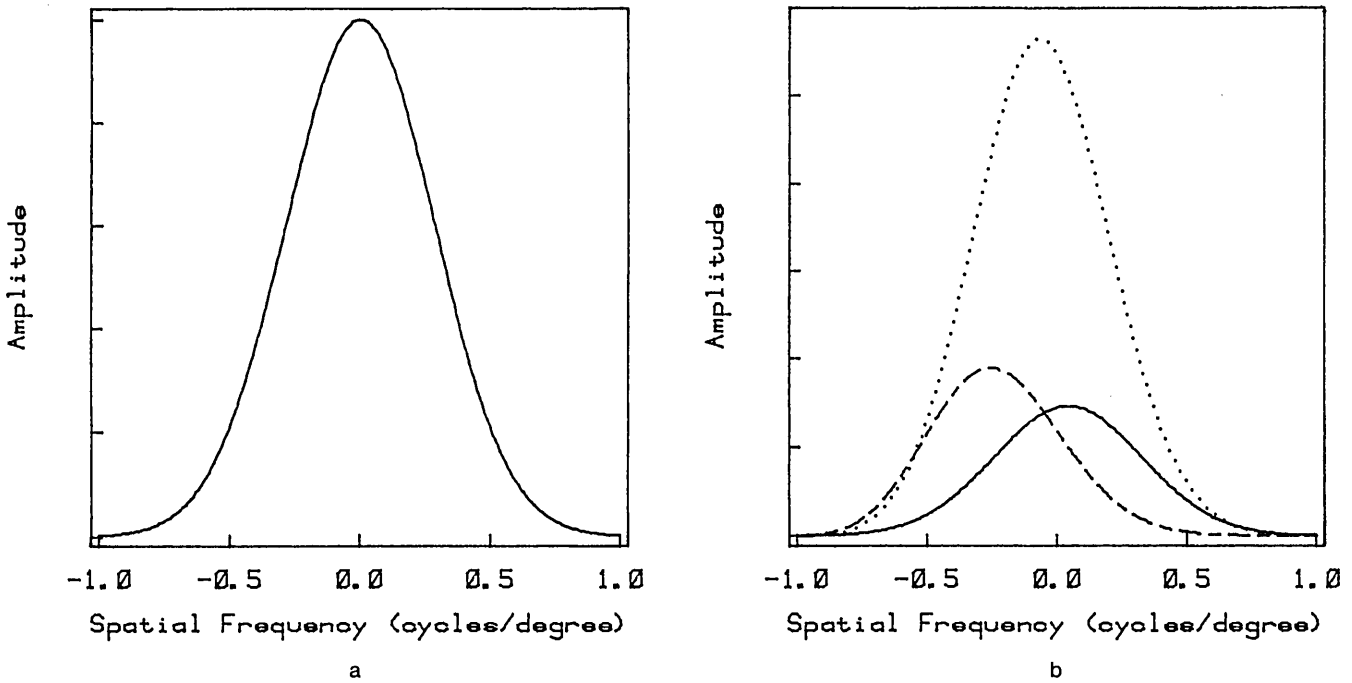


Fig. 11. Sensor-response amplitude as a function of the spatial frequency of a moving sinusoidal grating. a, Constant temporal frequency. b, Constant velocity. The shape of the curve is determined by the product of speed and sensor spatial frequency. The three curves are for values of 1 (solid), 16 (dotted), and 32 Hz (dashed).

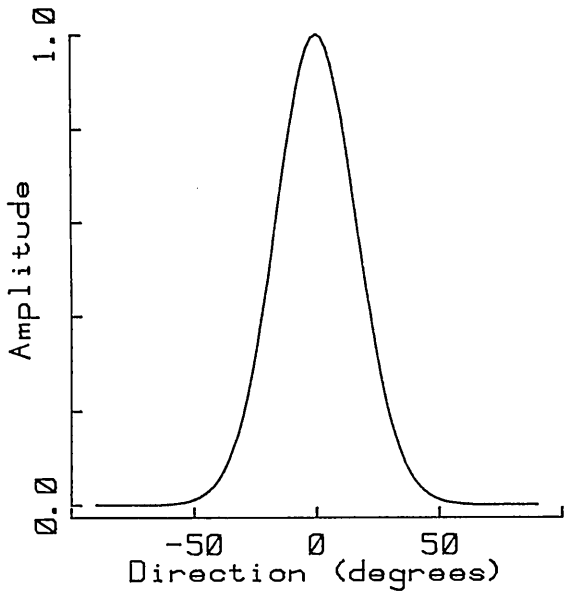


Fig. 12. Normalized response amplitude of the scalar motion sensor as a function of the direction of a moving sinusoid [Eq. (38)] with $\rho = 0.795$.

varies as a Gaussian function of the distance between grating and sensor frequencies. When $\lambda = 0.795/f$, as assumed here, the Gaussian has a width of one octave ($2f/3$) at half height. If the velocity is held constant, then $|\tilde{f}(rf)|$ will also vary, though not as rapidly as the Gaussian. These two cases are slices through the 3D amplitude response shown in Fig. 10b. They are illustrated in Fig. 11. The figure shows that at very low or very high speeds, the spatial-frequency band of the sensor can be shifted somewhat from its nominal center frequency.

2. Direction

The sensitivity of the sensor to the direction (and orientation) of a grating input can be determined by setting the spatial frequency of the grating to that of the sensor ($f = f_s$). Then the sensor response amplitude as a function of grating direction will be

$$G \exp(-\pi^2 s^2[f_s^2 + f^2 - 2f_s f \cos(\theta - \theta_s)])|\tilde{f}(rf)|. \quad (38)$$

The variation with orientation is captured in the exponential term, which simplifies to

$$\exp\left\{-\left[2\pi\rho \sin\left(\frac{\theta - \theta_s}{2}\right)\right]^2\right\}. \quad (39)$$

This function is drawn in Fig. 12. Making use of the approximation $\sin(x) \approx x$ for small x, we see this is approximately a Gaussian function of the difference in direction between the sensor and the grating:

$$\approx \exp\{-[\pi\rho(\theta - \theta_s)]^2\}. \quad (40)$$

This function is not visually discriminable from Eq. (38) shown in Fig. 12. If $\rho = 0.795$, as has been assumed here, then the sensor has an direction bandwidth of about 38°.

4. VECTOR MOTION SENSOR

A number of the results cited in Section 1, particularly those that deal with events near detection threshold, suggest that each direction is served by an independent sensor, such as the scalar sensor developed in Section 3. Other results, however, particularly those that involve estimating the apparent speed and direction of a superthreshold pattern, indicate the cooperative action of a number of sensors tuned to different directions.

Furthermore, we have seen that the output produced by an

individual scalar sensor is ambiguous with regard to the speed and direction of the image. Here, we show that this ambiguity can be resolved by combining the responses of a number of sensors. This combination is done by the second stage of the model, which we call the vector motion sensor. Before we build the vector sensor, however, we describe how the individual scalar sensors are distributed over space, spatial frequency, and orientation.

A. Distribution of Sensors

We have developed a model of a single sensor tuned for a particular location x_s, y_s, spatial frequency f_s, and direction θ_s. The full model contains many sensors, replicated over space, frequency, and orientation according to the following rules, which are simplified from an earlier model of spatial sensing.[46]

1. *Spatial Frequency*

As noted in Section 3, we have assumed that each sensor has a half-amplitude spatial-frequency bandwidth of one octave. This constrains the choice of the number of separate frequency bands to be analyzed, since too many will be redundant, whereas too few will result in a loss of information. We adopt an interval of one octave between bands. To cover the range of visible frequencies, we assume a set of eight sensor center frequencies of 1/4, 1/2, 1, 2, 4, 8, 16, and 32 cycles/deg. Further investigation may reveal whether all these bands are necessary or whether, for example, motion analysis is confined to lower spatial frequencies.

2. *Direction*

As noted in Section 3, the orientation bandwidth of each sensor is about 38°. To cover the full range of directions at intervals of about one bandwidth, we therefore assume 10 different possible sensor directions, moving in steps of 36° from 0°. These 10 directions can be constructed from the five orientations and two phases assumed in the spatial model described by Watson.[46]

3. *Spatial Sampling*

Loss of information through aliasing will be prevented if the sampling density of the sensors (the inverse of the distance between adjacent sensors) is at least twice the full-amplitude bandwidth of the sensor. We approximate this condition by use of a sampling density of four times the center frequency of the sensor (six times the half-amplitude bandwidth). With this density, the sampling artifact and the original spectrum overlap at a negligible amplitude. The actual sampling density required to account for human performance is a subject for further study.

Since the sampling density is proportional to frequency, there will be many more sensors at high spatial frequencies than at low. For example, in one square degree of visual field there will be one sensor of the lowest frequency (1/4 cycle/deg) but 16,384 sensors of the highest frequency (32 cycles/deg).

There are several aspects of the way in which sensors are likely to be arranged in the human visual system that we have not yet tried to include in our model. Most prominent is the way in which sensors may change in size with distance from the fovea. Second is the approximate hexagonal packing of sensors across the visual field. Both features are included in the model on which the spatial features of this model are based, but for simplicity we assume here a homogeneous square sampling array.

B. Direction Group

As noted, each scalar sensor provides ambiguous information about the image velocity, but this ambiguity can be resolved by combining scalar sensor responses. This combination is done within each set of 10 scalar sensors at the same location x_s, y_s and spatial frequency f_s that differ in direction θ_s. We call each such collection *a direction group*. Consider an image moving over this group. If the image contains many different orientations at frequency f_s, then many members of the group will respond. Since each corresponds to a different direction, no single one can indicate the actual direction in which the image moves. How do we deduce this actual direction from the responses of the group?

The response of any one scalar sensor does not indicate the velocity because of the direction ambiguity problem discussed earlier. Each scalar sensor conveys only one component of the image-velocity vector—the component in the sensor direction. But the responses of several sensors in the direction group provide several linearly independent components of the velocity vector. Therefore the full velocity vector can be determined from several (ideally, just two) scalar-sensor responses. This conversion of the scalar-sensor responses of the direction group to a single estimate of image velocity is done by the vector sensor.

Recall that the image-velocity components are coded in the temporal frequency of the scalar-sensor responses. If the image moves with velocity $\mathbf{r}$, then the temporal frequency in a sensor of frequency $\mathbf{s}$ will be $w_s = \mathbf{r} \cdot \mathbf{s}$. Converting this to polar terms, we see that the pattern of temporal frequencies among the sensors within the direction group is

$$w = r f_s \cos(\theta - \theta_s), \tag{41}$$

where (r, θ) are the speed and direction of the image and (f_s, θ_s) are the spatial frequency and direction of the sensor. This function describing the temporal frequency of the response of the sensor is a cosine with a phase equal to the actual direction of image motion and with an amplitude equal to the product of sensor frequency and image speed. Only half of the cosine can be estimated, since for half of the possible directions the sensor will be silent (the sensor responds only if its direction is within 90° of the direction of image motion).

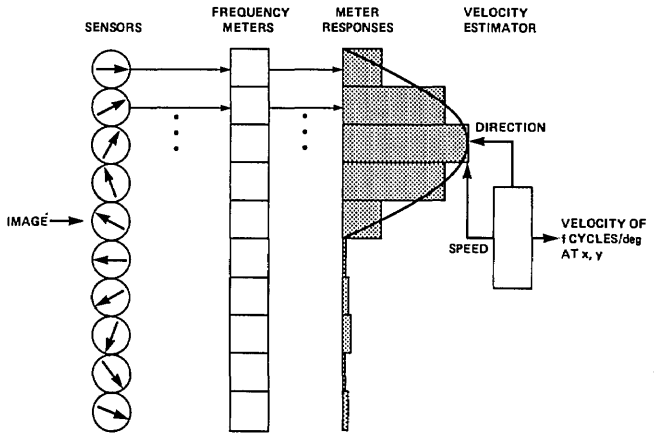


Fig. 13. The structure of the vector motion sensor.

There are a number of possible ways to extract the image velocity from this pattern of responses. The one that we have pursued is first to measure the temporal frequency of each scalar-sensor response (by means of a frequency meter) and then to fit Eq. (41) to the resulting quantities. This sequence is illustrated in a process diagram of the vector motion sensor in Fig. 13.

C. Inhibition between Opposite Directions

When an image moves past a direction group, it will excite half of the sensors in the group. This is because each sensor has a direction range of 180°. However, noise may generate responses in the remaining half of the sensors. We know, however, that only one of the pair of sensors tuned for opposite directions should respond to an image in motion. Thus we can null the noise responses by making each sensor compete with its opposite number. The larger of the two responses is preserved; the smaller is discarded. In fact, to make things simpler, we replace the smaller with the negative of the larger. The result is that the temporal frequency as a function of sensor direction is now a complete cycle of a cosine, rather than just a half cycle.

D. Summary

The result of all of the preceding operations will be a set of eight sampled vector fields, one per scale or spatial-frequency band. Each sample will be a vector indicating the velocity (speed and direction) of the image Gabor component at that location. The density of samples in each field will be proportional to the square of the spatial frequency. In Section 5, we describe some specific details of how these vector fields may be computed. Although we do not attempt at this time to provide a specific linking hypothesis between these vectors and perception, some common-sense predictions are possible. For example, if the vectors at all scales and locations agree, we predict that the image will appear to move with the specified velocity.

5. IMPLEMENTATION OF THE MODEL

A. General

The current implementation is written in the language *c* under the UNIX operating system. It currently resides on a SUN computer (an MC68010-based virtual-memory microcomputer with dedicated graphics display). The software is portable to other UNIX installations and will be described in a subsequent report.

B. Terminology

The input to the model is a continuous distribution of luminance or of contrast over the three dimensions of x, y, and t. A discrete representation of this input can be constructed by Nyquist sampling, resulting in what we will call a *movie*, a 3D array of values with dimensions W (width), H (height), and L (length). The values in the array may be either real or complex. A movie may be used to represent a sequence of images, its 3D Fourier transform, or any data set that fits within this structure. When used to describe a discrete moving image, the natural units of the movie are pixels and frames (or widths and lengths). When used to describe the transform of a discrete moving image, convenient units are cycles/width and cycles/length.

C. Scale

The input is analyzed in parallel by eight sets of sensors, each set selective for a different spatial-frequency band. The action of each set is identical to that of every other except for a change of spatial scale. This scale invariance allows us to describe the action of the sensors at just one scale k.

Consider an input digitized to a movie of size $W \times H \times L$. In the following discussion, we assume for simplicity that $W = H$. The digital movie contains spatial frequencies up to the Nyquist limit of 2^{-1} W cycles/width. The highest-sensor-frequency band considered will be centered at 2^{-2} W cycles/width. For convenience, we assign a scale number of 0 to this highest-frequency set of sensors applied to a given input. Since the sensors change spatial frequency in octave steps, the scale-k sensors will have a spatial-frequency passband centered at $2^{-(k+2)}$ W cycles/width.

Scaled Inputs The band-limited character of the sensors at each scale also permits an important computational economy. As noted, the input contains spatial frequencies up to $W\,2^{-1}$. The scale-0 sensor has a center frequency of $W\,2^{-2}$ and an effective passband cutoff one octave about this point, that is, $W\,2^{-1}$, equal to the cutoff frequency of the input [in fact, this is a very conservative figure, as the sensor frequency response has fallen by about 99.8% (54 dB) at this frequency]. At scale 1, the cutoff frequency of the sensor is reduced by a factor of 2 to $W\,2^{-2}$. Frequencies above this point may be removed from the input without altering the response. After these frequencies are removed, the number of samples required to define the input may be reduced by a factor of 2 in each spatial dimension. We define the *shrink* operator (denoted *Sh*) as one that removes the frequencies above one half of the Nyquist limit and then reduces the number of samples in each spatial dimension by a factor of 2. Our shrink algorithm is described in Fig. 14.[47–49] We define $\mathbf{C}_k$, the scale-k

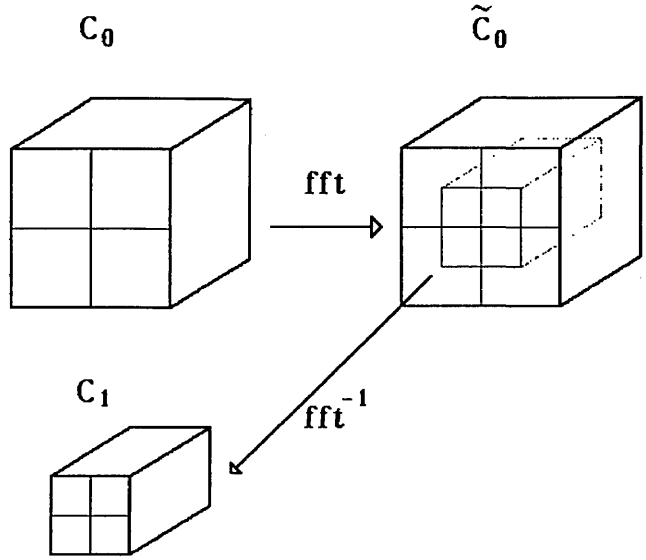


Fig. 14. The shrink algorithm. The original image of width W is denoted $\mathbf{C}_0$. Application of a fft results in $\tilde{\mathbf{C}}_0$. It contains frequencies up to 2^{-1} W. A square core, which contains frequencies up to 2^{-2} W, is inverse transformed to yield a movie $\mathbf{C}_1$ that is half as large in each spatial dimension. The procedure can be repeated to yield $\mathbf{C}_2$, and so on. The algorithm can be generalized to other size ratios and to the time dimension. This procedure is similar to other pyramid schemes[47–49] but has the advantage of precisely band limiting the signal before subsampling.

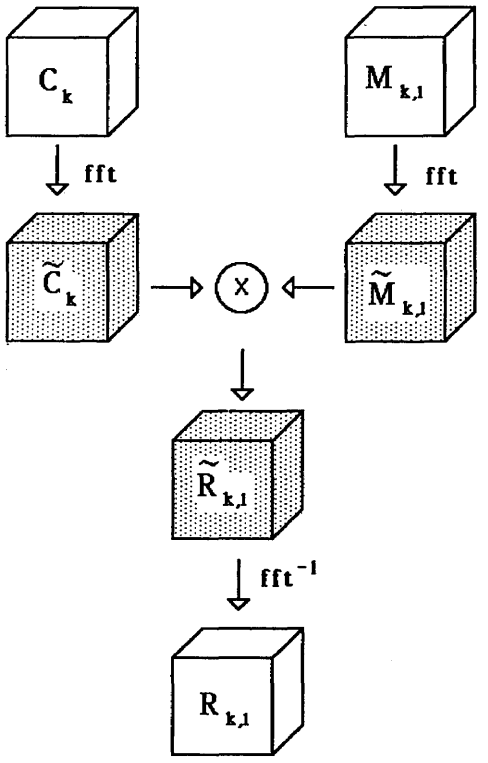


Fig. 15. Computation of $\mathbf{R}_{k,l}$, the sensor response at scale k, direction l. Shaded objects are Fourier transforms.

representation of the input, to be the result of applying the shrink operator to $\mathbf{C}_{k-1}$, and we write this as

$$\mathbf{C}_k \xrightarrow{Sh} \mathbf{C}_{k+1}. \tag{42}$$

The scale-0 representation $\mathbf{C}_0$ is the original input. Since the number of samples at each scale is reduced by a factor of 4 from that at the previous scale, the computational cost of repeating operations at multiple scales is greatly reduced. In fact, this multiple-scale representation of the input is less than 4/3 as large as the original input. Note also that since most processing occurs in the frequency domain, the shrink operation requires essentially no additional computation.

D. Input

The input is given as a movie $\mathbf{C}_0$ of normalized real floating-point values, with dimensions $W \times H \times L$. The movie dimensions can be mapped to world coordinates (degrees of visual angle, seconds) by a choice of scale factors. In the examples shown, the spatial scale was left undefined, but the time scale was set at 80 frames/sec.

E. Summary of Linear Operations

The first stage of the model is a set of linear sensors that differ in scale (spatial frequency), direction, and location. The response of a sensor of a particular direction, scale, and location is determined by cross correlating the sensor weighting function with the input. The responses of all sensors of a particular direction and scale are computed by convolving the sensor impulse response and the input and by sampling the result at the sensor locations. The convolution is done by transforming the impulse response and the input to the frequency domain, multiplying, and inverse transforming. Each spatial sample in the result is the time waveform of the response of the sensor at one location. This sequence of operations for scale k and direction l is pictured in Fig. 15.

1. *Sensor Impulse Response*

The impulse response of a sensor at scale 0 and direction l is defined by a real movie $M_{0,l}$ of the same width, height, and length as the input. The center spatial frequency of the sensor is $W\,2^{-2}$ cycles/width, where W is the input width in pixels. For example, if the input is a movie of width 32, height 32, and length 16, then the sensor impulse response at scale 0 will be the same size and will have a center frequency of 8 cycles/width. Values of the impulse response are sampled from Eq. (31). Figure 11a shows an example of an impulse response.

2. *Sensor Transfer Function*

To obtain the transfer function of the sensors at scale 0, the scale-0 impulse response is transformed:

$$M_{0,l} \xrightarrow{\text{fft}} \tilde{M}_{0,l}, \tag{43}$$

where fft indicates a fast Fourier transform. The result is a complex movie $\tilde{M}_{0,l}$ of size $W \times H \times L$ that defines the frequency response of the sensor up to $W\,2^{-1}$ cycles/width in both u and v and 40 Hz in w. Figure 11b shows an example of the magnitude of the transfer function of a sensor.

3. *Scaled Transfer Functions*

The scale-k sensor transfer function is equal to the scale-$k-1$ transfer function minified by a factor of 2 in u and v dimensions. It can be obtained by subsampling the transform of the previous scale in u and v dimensions. We denote this subsampling operation by the symbol Su (Fig. 16). In this way, the sensor transfer functions for the various scales are easily obtained from the transfer function at scale 0:

$$\mathbf{M}_{k,l} \xrightarrow{Su} \mathbf{M}_{k+1,l}. \tag{44}$$

4. *Sensor Response*

The input is Fourier transformed,

$$\mathbf{C}_k \xrightarrow{\text{fft}} \tilde{\mathbf{C}}_k, \tag{45}$$

yielding a complex movie of size $W\,2^{-k} \times H\,2^{-k} \times L$. The input and impulse response transforms are scalar multiplied,

$$\tilde{\mathbf{M}}_{k,l}\,\tilde{\mathbf{C}}_k = \tilde{\mathbf{R}}_{k,l}, \tag{46}$$

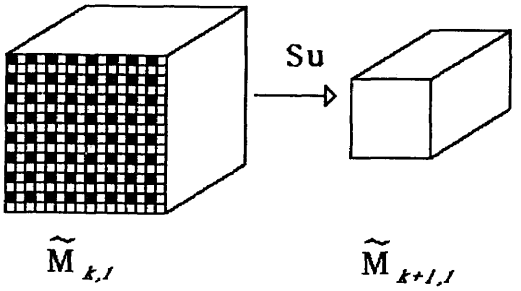


Fig. 16. Illustration of the frequency-subsampling operator Su. The array at each successive scale is obtained by sampling every other element in u and v dimensions.

and the result is inverse transformed, yielding the response

$$\tilde{\mathbf{R}}_{k,l} \xrightarrow{\text{fft}^{-1}} \mathbf{R}_{k,l}. \tag{47}$$

The value at x, y, t in this real movie specifies the response of a sensor of scale k and direction l at location x, y sampled at time t. (An example is pictured in Fig. 20, below.)

5. Border Effects

The method of convolution used above implicitly assumes that the input is periodic in x and y. Hence the outputs at the borders are due in part to input from the opposite border. These invalid results must be removed by stripping away a border $(N-1)/2$ elements wide from the output, where N is the width of the nonzero extent (support) of the impulse response. Here N is approximately seven pixels independent of scale, so we remove a border three pixels wide from the output of each convolution (an alternative would be to ensure that each input had a null border at least three pixels wide).

F. Summary of Nonlinear Operations

The operations performed in the second stage of the model are nonlinear. They combine responses of sensors within a velocity group, which up to this point have been kept separate. (Recall that a velocity group is the set of sensors of the same scale and location that differ in direction.) The principal steps are

(1) Measure the temporal-frequency spectrum of each sensor response.

(2) Select the temporal frequency with the largest magnitude.

(3) Find which of each pair of opposite-direction sensors (within the same velocity group) has the larger magnitude, save its frequency and magnitude.

(4) Within each velocity group, apply a fft to the vector of temporal-frequency values, considered as a function of direction.

(5) Use the amplitude and the phase of the first harmonic in the digital Fourier transform to calculate speed and direction, respectively.

(6) Save the largest magnitude within each velocity group as a measure of the strength of the response. If this magnitude is less than a threshold, assign the velocity "undefined" to the sensor.

G. Measuring Temporal Frequency

Since the spatial-frequency spectrum of the sensor is a Gaussian, it seems likely that the temporal-frequency spectrum of the response to a broadband stimulus would be roughly Gaussian. The desired quantity would be the location of this Gaussian. As a simple approximation to this, we have determined the frequency at which the largest magnitude occurs. Thus at each scale, direction, and location, a fft is applied to the vector of time samples of the sensor response. The frequency with the largest response magnitude and the magnitude itself are saved. The result at this stage is, for each scale k and for each direction, a movie of size $W_k \times H_k \times 2$.

1. Combining Responses within the Direction Group

In theory, only half of the sensors should respond to a moving stimulus: Those with directions more than 90 deg away from the direction of image motion should be silent. But since in the human visual system and in this implementation there are both noise and approximation, the sensors in the null direction may not be truly silent. Therefore we use a sort of inhibition to cancel their responses. Within each direction group, we compare the magnitudes of each pair of oppositely directed sensors. The larger is preserved; the smaller is replaced with minus the larger. We do this rather than null the smaller only to simplify the computations of Subsection 5.G.1. Finally, we compare all the magnitudes within the direction group and preserve the largest. We will call this the *strength* of the direction-group response.

The result at this step is, for each location and scale, a movie of size $W_k \times H_k \times (D+1)$, where D is the number of directions (10). The array specifies, for each location, the peak frequency as a function of direction (D values) and the strength of the velocity-group response.

2. Estimating Speed and Direction

The frequency-versus-direction function at each location should be a single cycle of a cosine with an amplitude equal to the product of the sensor spatial frequency and the speed of image-component motion and with a phase equal to the direction of image-component motion. The amplitude and the phase are computed as the first-harmonic results of a fft on the frequency-versus-direction vector. The final output for each scale is then a movie of size $W_k \times H_k \times 3$. The first and second frames contain speed and direction estimates, respectively, and the last frame contains the strength estimates. As noted in Subsection 5.E.5, a border of width 3 has been assigned a strength of 0.

6. RESULTS

We have applied our computational model to a small number of test cases. The inputs were either digitally synthesized or

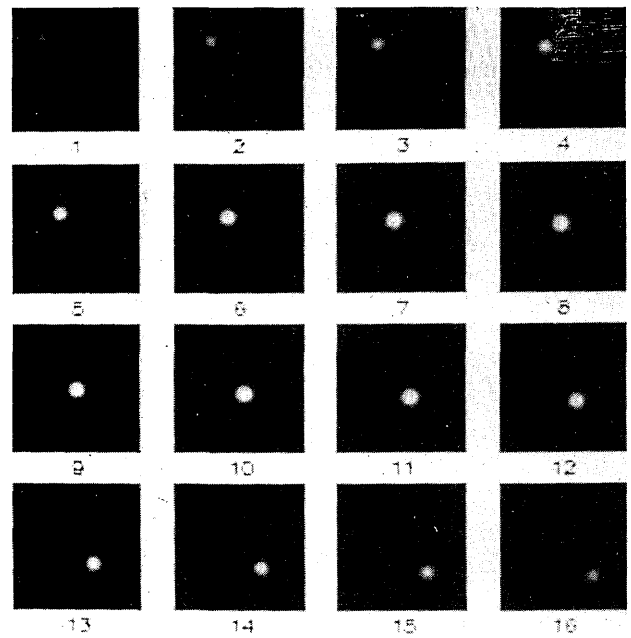


Fig. 17. 3D Gaussian blob. The dimensions are $W = 32, H = 32, L = 16$. The speed is $\sqrt{2}$ pixels/frame, and the direction is 315°. The spatial spread is two pixels; the temporal spread is eight frames.

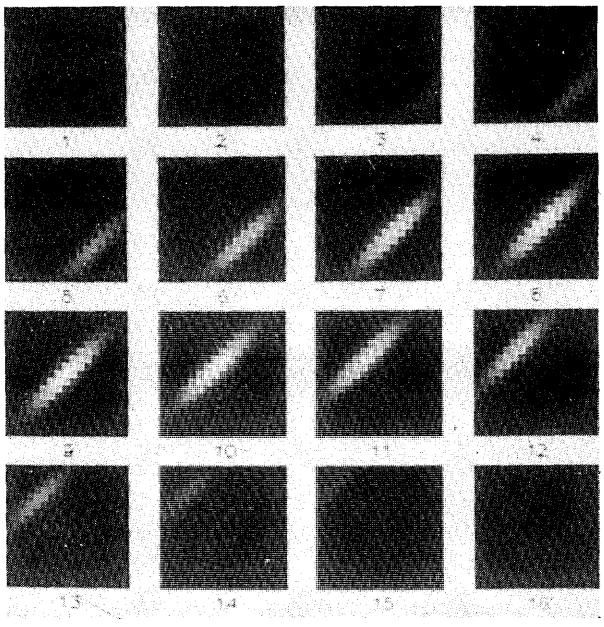


Fig. 18. Amplitude spectrum of the Gaussian blob. The axes are as in Fig. 11b.

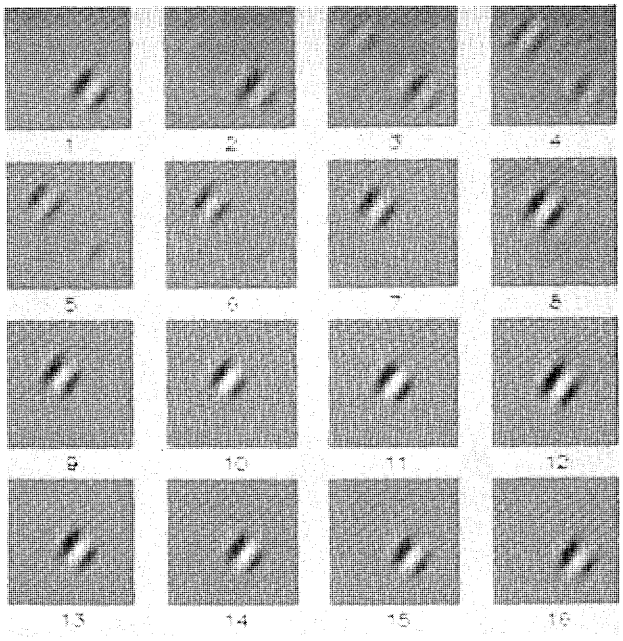


Fig. 19. Response of the scalar sensors of frequency 4 cycles/width and direction −36°.

digitized from a television camera. In both cases we have kept the spatial dimension small to minimize execution time. The length of each input was 16 frames (200 msec).

A. Three-Dimensional Gaussian Blob

The Gaussian blob is shown in Fig. 17. It is a 2D spatial Gaussian whose contrast varies as a Gaussian function of time as it traverses the field from upper left to bottom right. This is a useful input, because it contains all spatial frequencies and all move in the same clear direction. It also can be adequately sampled at the low resolution we are using.

It is instructive to look at the amplitude spectrum of this input (Fig. 18). Note that it consists of a Gaussian disk (broad in u and v, narrow in w) that has been sheared into a plane at 45° to both the u and the v axes. The energy in the fourth quadrant exists only for negative time frequencies (early frames in Fig. 18), while energy exists in the second quadrant only for positive frequencies. This should be compared with the amplitude response of the sensor pictured in Fig. 11b.

It is also of interest to examine the response of the scalar sensors of one directed frequency (scale and direction) to the Gaussian blob input. This is shown in Fig. 19. The response

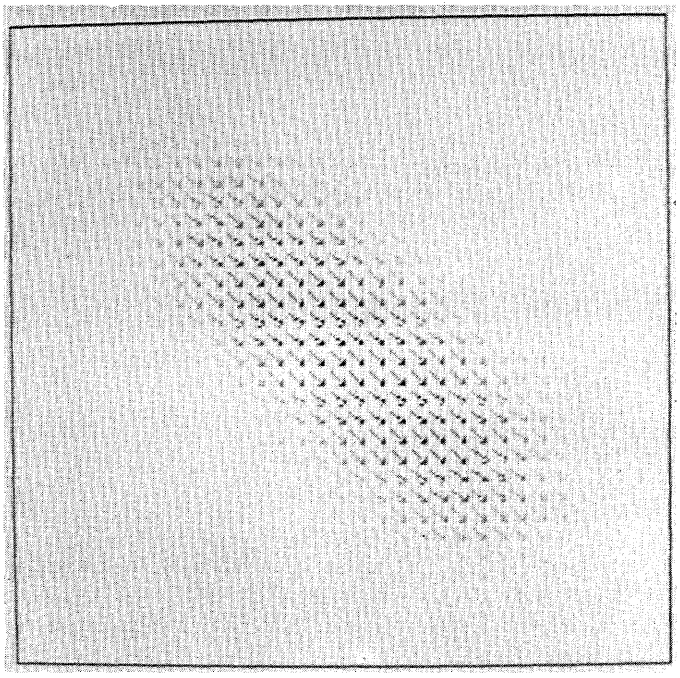


Fig. 20. Response of the vector sensors at scale 0 to the Gaussian blob. Each arrow is an estimate of image velocity at the corresponding spatial frequency and location. The contrast of the arrow indicates the strength of the response.

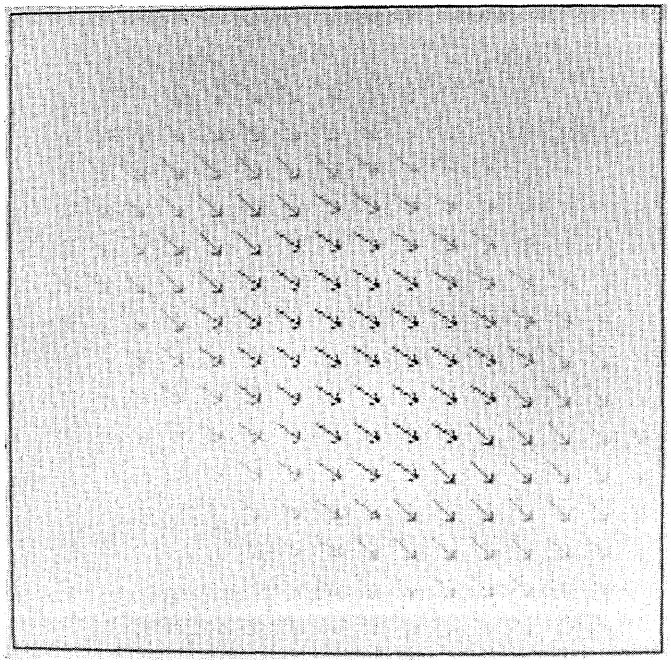


Fig. 21. Vector-sensor responses to the Gaussian blob at scale 1.

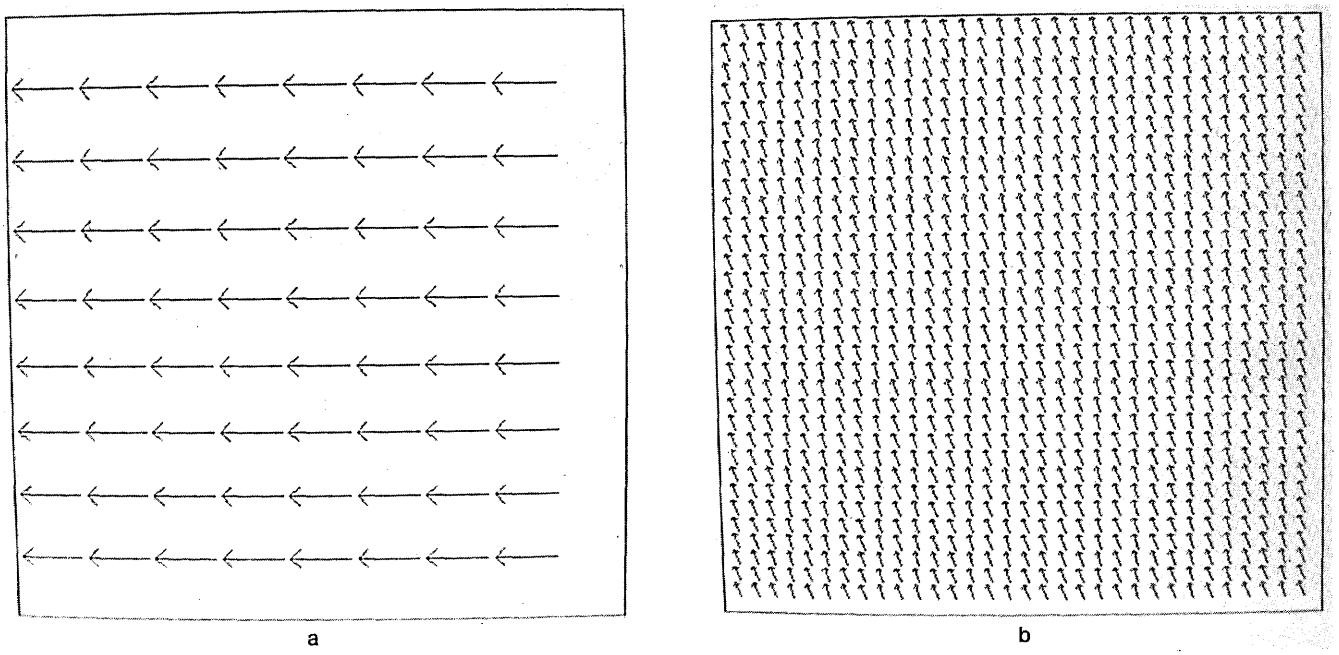


Fig. 22. Simulated responses to the sum of two gratings of different spatial frequencies that move in different directions. The frequencies were 2 and 8 cycles/width, the directions were 90° and 180°, and the speeds were both 1 pixel/frame. a, Response at the scale of the lower frequency. b, Response at the scale of the higher frequency.

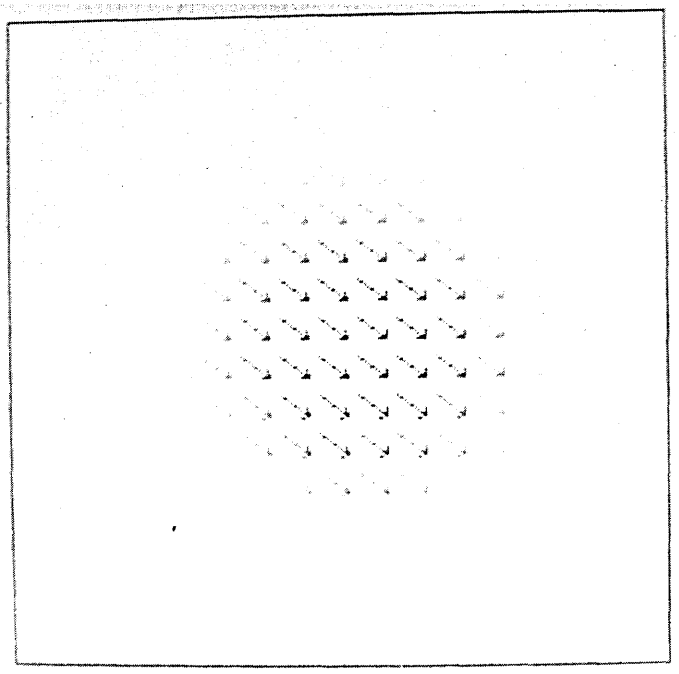


Fig. 23. Simulation of apparent motion. The input was as shown in Fig. 17 with all but frames 6 and 8 blank. The output at a scale of 4 cycles/width is shown.

takes some time to develop and makes its first substantial appearance in the upper-left-hand corner of frame 3. The responses in the lower right-hand corner of the first four frames are an artifact of the use of a circulant convolution method and should be interpreted as frames 17–20 of the response, wrapped around into the first four frames. Note that the sensor passes only spatial frequencies near that of the sensor, with the result that, for this input, the response looks much like the spatial weighting function of the sensor moved along the path of the input. For sensors of other directions, the result would look the same, except that the amplitude of the response and the orientation of the bars would be different.

Figure 20 shows the response of the vector sensors at scale 0. Since the input has a width of 32, this corresponds to a spatial frequency of 8 cycles/width, the highest frequency (smallest scale) that we can compute on this input. Each arrow represents the response of one vector sensor. The speed is indicated by the length of the arrow and the direction by the angle of the arrow. In addition, the intensity of the arrow is proportional to the strength of the sensor response. Arrows have not been plotted whose intensity or length is less than 10% of the maximum within the picture.

The arrows point in the correct direction and have approximately the correct length. Furthermore, strong responses occur only along the actual path of motion. For this input, at this scale, the model gives a good indication of the true and perceived velocity of image components.

Figure 21 shows the response at the next scale (4 cycles/width). The results are similar. There are fewer arrows because the sampling density of the sensors is lower at this scale. The responses are less confined to the path of motion, as we would expect of sensors with lower spatial resolution.

B. Coherent Gratings

We have simulated the experiment of Adelson and Movshon described in Subsection 1.D. We first superimposed two gratings with the same spatial frequency (8 cycles/width) moving at 1 pixel/frame in directions of 90° and 180°. To the human observer, this appears as a coherent plaid that moves to the upper left (135°). The model response at a frequency of 8 cycles/width, where the sensors are matched to the spatial frequency of the stimulus, correctly indicated motion at about the right speed in about the right direction.

We then added together a grating of 8 cycles/width and

direction of 90° and one of 2 cycles/width and direction 180°, both with a speed of 1 pixel/frame. This appears to the human as a pair of separate gratings gliding over each other, each moving in its own direction. As shown in Fig. 22, the model also indicates that each frequency should appear to move in its own direction.

C. Apparent Motion

The input was a pair of Gaussian spatial blobs, as described above, that appeared only in frames 6 and 8 of the 16-frame sequence. The distance between blob centers was two pixels

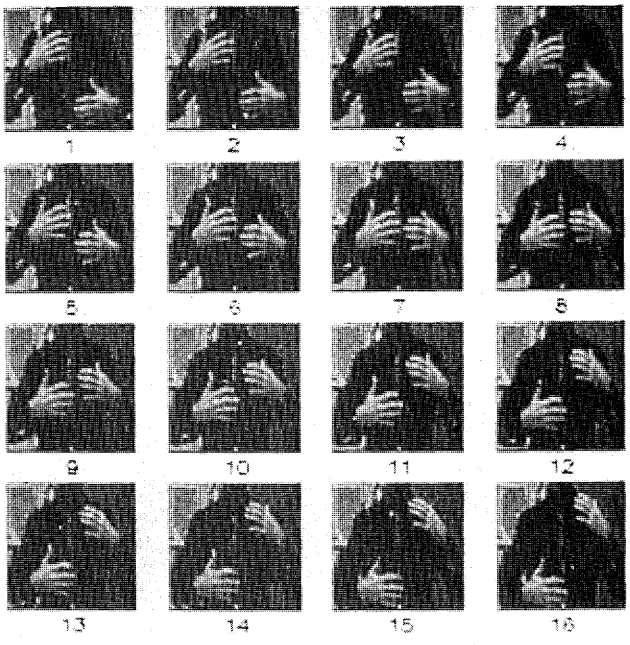


Fig. 24. A sequence of natural images in which two objects (the hands) move in different directions. The width is 32 pixels.

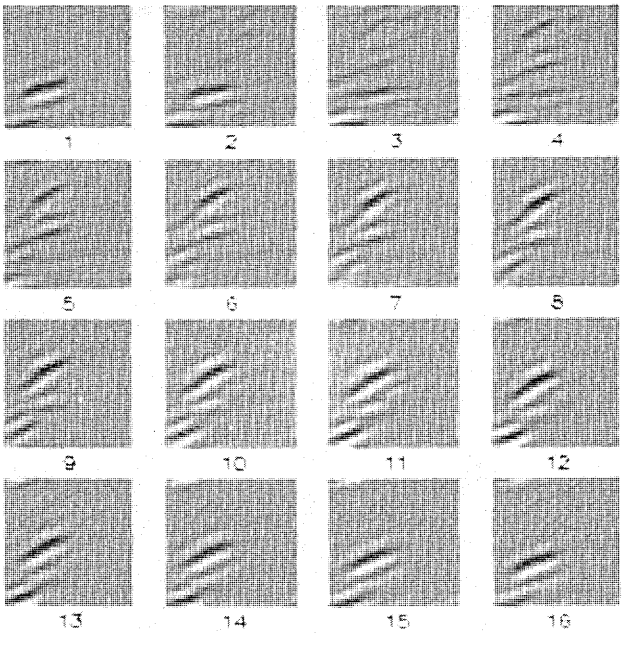


Fig. 25. The scalar-sensor responses $\mathbf{R}_{k,l}$ to the input in Fig. 24. Scale, 8 cycles/width, direction, 108°. The first four frames are wrap-arounds from the end of the response.

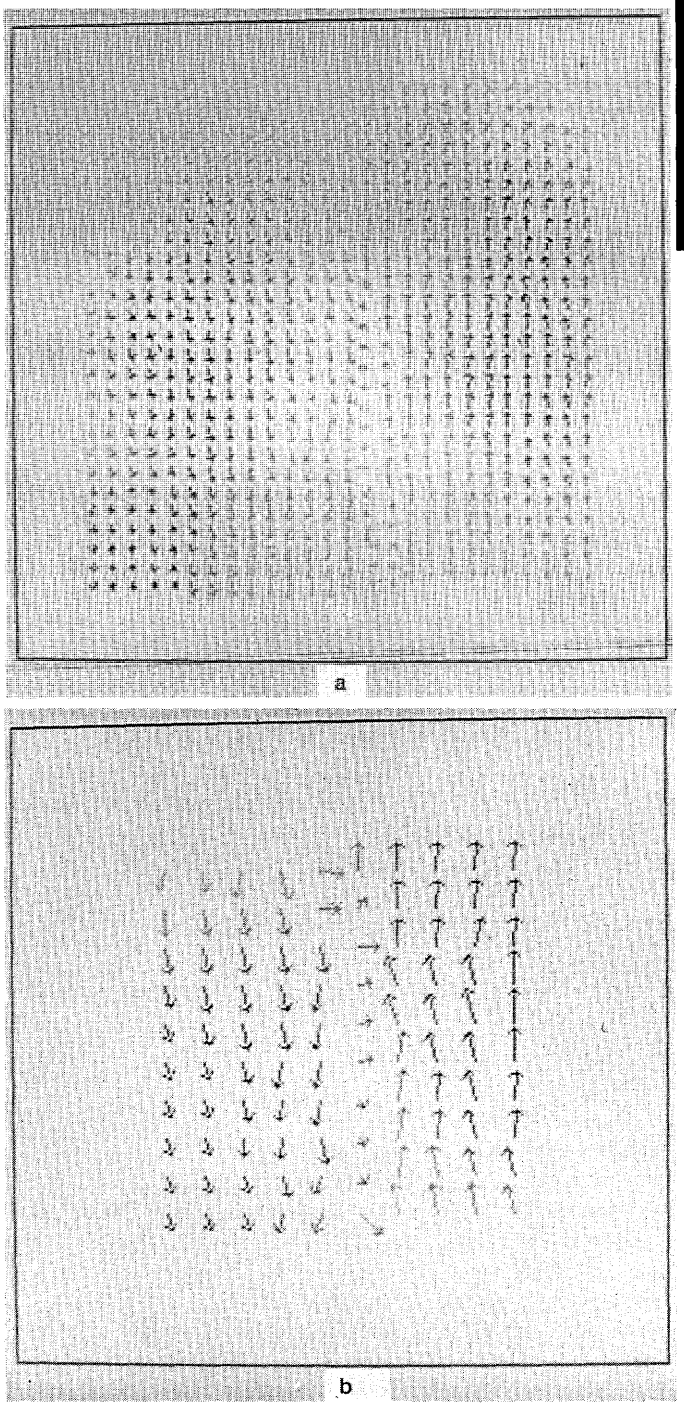


Fig. 26. Vector-sensor responses to the hand-waving sequence. a, Output at a scale of 8 cycles/width. b, Output at a scale of 4 cycles/width.

in each of x and y. This is the same as the sequence in Fig. 17 with all but frames 6 and 8 zeroed. At the appropriate spatial scale, this looks to the human like a blob that moves from the first point to the second. The vector-sensor output is shown in Fig. 23. Qualitatively at least, the model correctly predicts the preceived motion.

D. Hand Waving

Our final example is a natural image sequence digitized from a video camera (Fig. 24). It shows a person moving his hands.

It is an interesting case because it contains objects moving in different directions at different places. It is edifying to examine the responses of the scalar sensor for a particular scale and direction. Figure 25 shows the scalar response $\mathbf{R}_{k,l}$ at a scale of 8 cycles/width in a direction of 108°. These sensors respond only to the upward-moving right hand, all other image components having been filtered out.

The vector responses are shown in Fig. 26. The output at both scales shows responses that are approximately correct as to speed and direction (as compared with the nominal speed and direction of the hand) and agree with our subjective impression of the movement of the hands.

7. DISCUSSION

A. Relation to Other Models

Since our first descriptions of the scalar motion sensor,[5,7] two descriptions of similar mechanisms have appeared.[50,51] Both models resemble ours in beginning with linear spatiotemporal filters selective for spatial frequency, but both differ from ours in important respects. In the van Santen–Sperling model, the outputs of opposite-direction sensors are multiplied and the result integrated over some interval. In the model of Adelson and Bergen, the energy in each sensor output is integrated over some interval. Both of these procedures, which we characterize as energy models, are fundamentally different from the frequency-demodulation operation used in our model. The energy models go to some effort to remove the temporal modulation of the sensor response, whereas we preserve it and note that it directly codes the image-velocity components. It is interesting to note that energy models will be quite susceptible to variations in the contrast of image components at different orientations and directions. Our model will, on the other hand, be immune to these effects, since the temporal frequency of the scalar-sensor response is largely independent of contrast. This advantage is similar to that enjoyed by FM radio broadcasts over AM.

The descriptions of the energy models do not yet go beyond the level of a single scalar sensor. Both are defined in only one spatial dimension and thus do not confront the problem of estimating the full velocity vector from the ambiguous scalar-sensor responses. Neither model specifies the particular spatial or temporal impulse response of their sensors. Finally, both energy models are of a single sensor and hence do not specify how the sensors are distributed over space or frequency. These differences prevent further comparisons with our model, but it seems likely that experimental tests might distinguish between our model and more completely specified versions of the energy models.

B. Remaining Questions

We have constructed a system for processing a dynamic visual input (a movie), which assigns velocities to components of the input. The output may be regarded as a sampled, multiple-scale velocity field. The velocity assignments made by the system resemble those made by human observers, though no strict tests of this correspondence have yet been made.

A number of issues have been left unresolved, and several known features of human vision have not been modeled. For example, though our temporal impulse response is matched to human sensitivity, no similar match has been made in the spatial domain. The spatial contrast-sensitivity function will affect the precise correspondence between model outputs and human judgments. Likewise, the space-variant resolution of human vision may have important effects on motion perception, but it is not included in this version of our model.

Another intriguing and unresolved issue is the time scale of velocity assignments. There are at least four questions:

(1) How much time is used to compute a velocity estimate?
(2) How often is a new estimate computed?
(3) Are the answers to the previous two questions the same at each scale?
(4) To what moment in time is the velocity assigned?

The first three questions are straightforward and might be resolved by experiment, but the fourth requires further explanation.

At issue is how information from different sources is related in the time dimension. Just as estimates of velocity are assigned to particular locations and scales, so are they assigned to particular times. Just as we might suspect that information at different scales is kept in registration by a *topographic* mapping, so information from different mental processing channels must be kept in *chronographic* registration. For instance, to catch a ball in flight the human must keep in registration a number of visual representations (e.g., velocity, shape, color) as well as a motor representation of the position of the hand.

This example also illustrates the special problems in perception that are due to the causality of processing in the time domain. Visual processing takes time and must therefore lag behind the stimulus events in the world. But, to catch the ball, one must abolish this lag through prediction and anticipation. Therefore we must suppose that, based on a mental clock, velocities are assigned to the past and used to predict the present and future.

As noted in the introduction, a model of human motion sensing must pass two tests. The first is whether it successfully mimics psychophysical data on sensitivities, discriminations, illusions, and appearances. It has passed some preliminary tests of this sort, as described in Section 6. The second test is how useful and efficient the representation is for further mental computations. This is likely to be answered only by trying to model higher visual functions, such as image segmentation, object recognition, and judgments of object and self-motion in four-dimensional space–time.

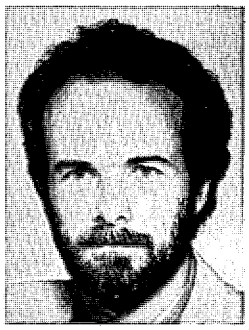
REFERENCES

1. J. J. Gibson, *The Perception of the Visual World* (Houghton Mifflin, Boston, Mass., 1950).
2. S. Ullman, *The Interpretation of Visual Motion* (MIT U. Press, Cambridge, Mass., 1979).
3. H. C. Longuet-Higgins and K. Prazdny, "The interpretation of moving retinal images," Proc. R. Soc. London Ser. B **208**, 385–387 (1980).
4. E. C. Hildreth, *The Measurement of Visual Motion* (MIT U. Press, Cambridge, Mass., 1983).
5. A. B. Watson and A. J. Ahumada, Jr., "A look at motion in the frequency domain," NASA Tech. Mem. 84352 (1983).
6. E. H. Adelson and J. A. Movshon, "Phenomenal coherence of moving visual patterns," Nature **300**, 523–525 (1982).

7. A. B. Watson and A. J. Ahumada, Jr., "A linear motion sensor," Perception **12,** A17 (1983).
8. F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," J. Phys. London **197,** 551–566 (1968).
9. N. Graham and J. Nachmias, "Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models," Vision Res. **11,** 251–259 (1971).
10. A. B. Watson, "Summation of grating patches indicates many types of detector at one retinal location," Vision Res. **22,** 17–25 (1982).
11. N. Graham, "Spatial-frequency channels in the human visual system: effects of luminance and pattern drift rate," Vision Res. **12,** 53–68 (1972).
12. A. B. Watson, P. G. Thompson, B. J. Murphy, and J. Nachmias, "Summation and discrimination of gratings moving in opposite directions," Vision Res. **20,** 341–347 (1980).
13. P. Thompson, "The coding of velocity of movement in the human visual system," Vision Res. **24,** 41–45 (1984).
14. S. P. McKee, "A local mechanism for differential velocity detection," Vision Res. **21,** 491–500 (1981).
15. A. Pantle, "Motion aftereffect magnitude as a measure of the spatiotemporal response properties of direction-sensitive analyzers," Vision Res. **14,** 1229–1236 (1974).
16. E. Levinson and R. Sekuler, "Adaptation alters perceived direction of motion," Vision Res. **16,** 779–781 (1976).
17. D. J. Tolhurst, "Separate channels for the analysis of the shape and the movement of a moving visual stimulus," J. Physiol. London **231,** 385–402 (1973).
18. E. Levinson and R. Sekuler, "The independence of channels in human vision selective for direction of movement," J. Physiol. London **250,** 347–366 (1975).
19. C. F. Stromeyer III, J. C. Madsen, S. Klein, and Y. Y. Zeevi, "Movement-selective mechanisms in human vision sensitive to high spatial frequencies," J. Opt. Soc. Am. **68,** 1002–1005 (1978).
20. R. J. W. Mansfield and J. Nachmias, "Perceived direction of motion under retinal image stabilization," Vision Res. **21,** 1423–1425 (1981).
21. J. G. Robson, "Spatial and temporal contrast-sensitivity functions of the visual system," J. Opt. Soc. Am. **56,** 1141–1142 (1966).
22. J. J. Koenderink and A. J. van Doorn, "Spatiotemporal contrast detection threshold surface is bimodal," Opt. Lett. **4,** 32–34 (1979).
23. D. H. Kelly, "Adaptation effects on spatio-temporal sine-wave thresholds," Vision Res. **12,** 89–101 (1972).
24. D. H. Kelly, "Motion and vision. II. Stabilized spatio-temporal theshold surface," J. Opt. Soc. Am. **69,** 1340–1349 (1979).
25. D. C. Burr and J. Ross, "Contrast sensitivity at high velocities," Vision Res. **22,** 479–484 (1982).
26. P. A. Kolers, *Aspects of Apparent Motion* (Pergamon, New York, 1972).
27. G. Sperling, "Movement perception in computer-driven visual displays," Behav. Res. Methods Instrum. **8,** 144–151 (1976).
28. A. B. Watson and A. J. Ahumada, Jr., "Sampling, filtering, and apparent motion," Perception **11,** A15 (1982).
29. A. B. Watson, A. J. Ahumada, Jr., and J. Farrell, "The window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays," NASA Tech. Pap. 2211, (1983).
30. M. Fahle and T. Poggio, "Visual hyperacuity: spatiotemporal interpolation in human vision," Proc. R. Soc. London Ser. B **213,** 451–477 (1981).
31. A. B. Watson and J. G. Robson, "Discrimination at threshold: labelled detectors in human vision," Vision Res. **21,** 1115–1122 (1981).
32. E. Levinson and R. Sekuler, "Inhibition and disinhibition of direction-specific mechanisms in human vision," Nature **254,** 692–694 (1975).
33. C. F. Stromeyer III, R. E. Kronauer, J. C. Madsen, and S. A. Klein, "Opponent-movement mechanisms in human vision," J. Opt. Soc. Am. A **1,** 876–884 (1984).
34. P. Thompson, "Discrimination of moving gratings at and above detection threshold," Vision Res. **23,** 1533–1538 (1983).
35. F. W. Campbell, J. Nachmias, and J. Jukes, "Spatial-frequency discrimination in human vision," J. Opt. Soc. Am. **60,** 555–559 (1970).
36. F. H. C. Crick, D. C. Marr, and T. Poggio, "An information processing approach to understanding the visual cortex," in *The Organization of the Cerebral Cortex*, S. G. Dennis, ed. (MIT U. Press, Cambridge, Mass., 1981).
37. J. G. Moik, "Digital processing of remotely sensed images," NASA Doc. SP-431 (1980).
38. The orientation of a 2D frequency component is the angle of a normal to the wavefront.
39. D. Marr and S. Ullman, "Directional selectivity and its use in early visual processing," Proc. R. Soc. London Ser. B **211,** 151–180 (1981).
40. F. L. van Nes, J. J. Koenderink, H. Nas, and M. A. Bouman, "Spatiotemporal modulation transfer in the human eye," J. Opt. Soc. Am. **57,** 1082–1088 (1967).
41. A. B. Watson, "Temporal Sensitivity," in *Handbook of Perception and Human Performance*, J. Thomas, ed. (Wiley, New York, to be published).
42. M. G. F. Fourtes and A. L. Hodgkin, "Changes in the time scale and sensitivity in the omatidia of limulus," J. Physiol. London **172,** 239–263 (1964).
43. A. B. Watson, "Derivation of the impulse response: comments on the method of Roufs and Blommaert," Vision Res. **22,** 1335–1337 (1982).
44. S. Marcelja, "Mathematical description of the responses of simple cortical cells," J. Opt. Soc. Am. **70,** 1297–1300 (1980).
45. B. Sakitt and H. B. Barlow, "A model for the economical encoding of the visual image in cerebral cortex," Biol. Cybern. **43,** 97–108 (1982).
46. A. B. Watson, "Detection and recognition of simple spatial forms," in *Physical and Biological Processing of Images*, A. C. Slade, ed. (Springer-Verlag, Berlin, 1983).
47. S. Tanimoto and T. Pavlidis, "A hierarchical data structure for picture processing," Comput. Graphics Image Process. **4,** 104–119 (1975).
48. P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. **COM-31,** 532–540 (1983).
49. J. L. Crowley and R. M. Stern, "Fast computation of the difference of low-pass transform," IEEE Trans. Pattern Anal. Mach. Intelligence **PAMI-6,** 212–222 (1984).
50. G. Sperling and J. P. H. van Santen, "Temporal covariance model of human motion perception," J. Opt. Soc. Am. A **1,** 451–473 (1984).
51. E. H. Adelson and J. R. Bergen, "Motion channels based on spatiotemporal energy," Invest. Ophthalmol. Vis. Sci. Suppl. **25,** 14 (A) (1984).

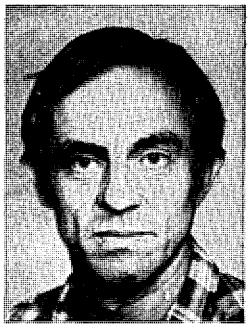
(*see overleaf*)

Andrew B. Watson



Andrew B. Watson did undergraduate work at Columbia University and in 1977 received the Ph.D. degree in psychology from the University of Pennsylvania. He spent three years in England at the Craik Laboratory of Cambridge University, followed by two years in the Department of Aeronautics and Astronautics at Stanford University. He has been at NASA Ames Research Center since 1982, where he heads the Perception and Cognition Group. Current research interests are computational models of visual perception and their application to image coding, analysis, and display.

Albert J. Ahumada, Jr.



Albert J. Ahumada, Jr., received the B.S. degree in mathematics from Stanford University in 1961 and the Ph.D. degree in psychology from the University of California at Los Angeles in 1967. He has taught at the University of California, Irvine, in the School of Social Sciences and done research in the Department of Aeronautics and Astronautics at Stanford University. He is now a research psychologist at NASA Ames Research Center, where his primary research interest is the development of mathematical models of visual processes.